MASTER THESIS

DATA LOCATION COMPLIANCE IN CLOUD COMPUTING

J. Noltes

MSC COMPUTER SCIENCE
TRACK INFORMATION SYSTEMS ENGINEERING

EXAMINATION COMMITTEE
dr. ir. W. (Wolter) Pieters
dr. ir. V. (Virginia) Nunes Leal Franqueira
M.J. (Mark) Butterhoff RE bc (KPMG)

DOCUMENT NUMBER
EEMCS - 0089990

UNIVERSITY OF TWENTE.

26-08-2011

Data location compliance in cloud computing

Master thesis
MSc Computer Science
Track Information Systems Engineering

Johan Noltes
s0089990
26 August 2011
Final version

Graduation committee
dr. ir. W. (Wolter) Pieters
dr. ir. V. (Virginia) Nunes Leal Franqueira
M.J. (Mark) Butterhoff RE bc (KPMG)

Management summary

The Gartner hype cycle defines 2011 as 'the year of the cloud'. Cloud computing combines the newest techniques to deliver new services, which are rapidly scalable, are using shared resources, offer pay-per-use and are delivered via a broadband network (e.g. internet). Consumers are rapidly adopting cloud computing, but business are hesitating. An important factor for this hesitation is that businesses need to be compliant to legislation.

An example of such legislation is the EU data protection directive, which states that privacy sensitive data should always be located within the European Union. However, due to the nature of cloud computing, the location of the data is often unknown, or may change frequently. Currently, Cloud Service Providers (CSPs) do not always offer services that comply to this data location legislation, or in case they do, they do not always show compliance to their customers. This research is about how CSPs can show compliance to customer demands regarding data location.

Interviews with CSPs show that CSPs are currently in principle able to determine and control the location of data of their customers, e.g. by using the configuration of the hypervisor. However, these CSPs do not give guarantees about the location of data.

This research proposes the Cloud Computing Compliance Guideline, based on interviews and literature study. The Cloud Computing Compliance Guideline gives a process description of showing compliance, which enables CSPs to show compliance to customer demands regarding data location. The Cloud Computing Compliance Guideline comprises of four phases.

Phase 1 describes how the customer prepares the movement to the cloud, by carrying out a risk assessment, data classification, creating security demands regarding data location and CSP selection. Phase 2 describes the negotiation process between the customer and CSP. The guideline describes two frameworks that can be used for the SLA negotiation: the SLA@SOI framework and the XACML framework. After the automated negotiation, the CSP takes security measures to ensure data will be stored conform the agreements. Phase 3 describes the regular storage process. Because all security measures are taken, no extra efforts are needed. However, the CSP monitors and logs the movement of data, to detect possible violations. Phase 4 describes how the CSP shows compliance to the customer demands regarding data location. This is done by regularly reporting the current status, and carrying out external audits to give assurance about the correctness of the process.

When these phases are carried out correctly, an auditor checks whether CSP executes the correct processes and data is stored on the allowed locations. If this is the case, the auditor can give assurance that the agreements with the customer are enforced, so the CSP can show compliance to the customer demands.

The Cloud Computing Compliance Guideline is validated using interviews with CSPs. These interviews indicate that CSPs think the Cloud Computing Compliance Guideline can be used in practice, but some adaptions are needed.

Preface

This document contains my master thesis, the final document that I produced for the master Computer Science at the University of Twente. It describes the results of my research on data location compliance in cloud computing, which I carried out at KPMG IT Advisory. During my period at KPMG, cloud computing became an important proposition for the company. I hope that the results of this research contribute to the knowledge and propositions within the company.

This master thesis would not have been possible without the support of many people, starting with my supervisors Wolter and Virginia. They helped me to get the right research approach, and continuously delivered high quality feedback. I would like to thank them for their guidance and support. In addition I would like to thank Mark, my supervisor at KPMG, for his guidance: his enthusiasm and quick reasoning helped me to make the right decisions after only a few questions.

I would also like to thank my fellow students in Enschede and my fellow colleagues at KPMG. They provided a pleasant atmosphere to work on this project, and the informal conversations brought me a lot of new insights, hints and feedback. But off course I am especially thankful for the great times we spent together and hopefully keep doing in the near future.

Finally, I would like to make a special note for my parents and family. They supported me throughout my entire study, and encouraged all great activities like my time at the board of Inter-*Actief* and the study tour to the United States. Thanks to their faith and support, I was able to finish my study and make it an great time to look back on.

I hope you will enjoy reading this master thesis about data location compliance in cloud computing. If you have any questions, please feel free to contact me.

Johan Noltes
Enschede / Amstelveen, August 2011

Contents

1 Introduction

According to the Gartner Hype Cycle [1], 2011 is 'the year of cloud computing'. In this year, many organizations are considering to start using the cloud. But what is cloud computing, is this hype something completely new? No. Since the start of professional IT use, the commoditization and centralization of IT has increased each year. Years ago, organizations had all their IT in their own server rooms, 'on premise'. Over the past years, the servers were shared with other businesses in shared service centers (SSC), while recently they have been outsourced to third parties. Cloud computing is the next central step in this evolution of IT, as depicted in Figure 1.

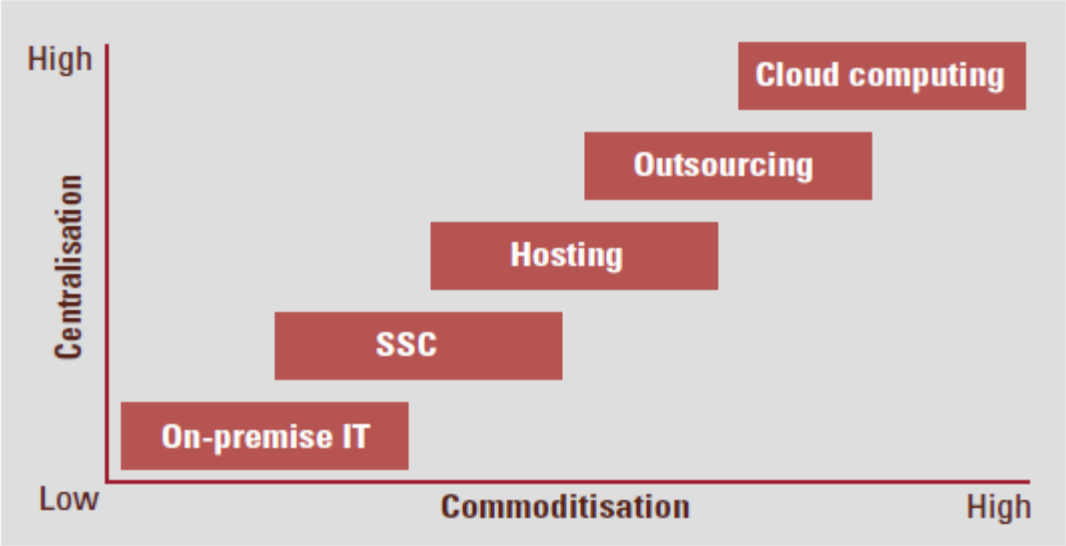


Figure 1 Paradigm shift in IT [2]

Cloud computing combines the newest techniques to deliver new services, which are rapidly scalable, are using shared resources, offer pay-per-use and are delivered via a broadband network (e.g. internet). Cloud computing can be offered in three service models which determine which components are offered by the CSP; Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). This can be done using four deployment models which determine with who resources are shared; private cloud, public cloud, community cloud and hybrid cloud. These concepts are described more elaborately in chapter 2.

1.1 Motivation

With the growing success of cloud computing, many organizations are considering to migrate their own applications and data into the cloud. The advantage of cloud computing is that the IT services offered are 'elastic'; customers only pay for the capacity used and can easily scale up or down, and do not have to make large investments in new hardware. Cloud computing leads to more flexibility, better scalability, higher availability, shorter time to market and better cost control [3]. This is especially important when the demand is unknown or when large peaks are expected. For example, a web startup needs to support a spike in demand when it becomes popular, followed potentially by a reduction once some visitors turn away [4].

1.1.1 Market situation

In 2010, KPMG held a survey with 125 respondents [5], all decision makers and business managers in the Netherlands. A 59% majority of them agrees with the statement that cloud computing is the future model of IT, while only 12% disagrees. The respondents believe that cloud computing is not a hype, but an important future IT concept.

The Gartner CIO Agenda 2011 [6] shows the results of a survey held with 2,014 CIOs. The respondents work across 27 industries and in 41 countries, and represent more than $159 billion in corporate and public sector IT spending. Cloud computing is ranked first as strategic technology priority for 2011, showing the importance CIOs attach to this technology.

Market-research firm IDC [7] expects IT cloud services spending to grow from about $16 billion in 2008 to about $42 billion by 2012 and to increase its share of overall IT spending from 4.2% to 8.5%. According to the research firm Gartner [8], global sales of cloud services rose 17% in 2010, to $68.3 billion from $58.6 billion in 2009. Global sales of cloud services are expected nearly to double by 2012, to $102.1 billion, Gartner estimates.

1.1.2 Risk

Despite the mentioned advantages and importance given to cloud computing by practitioners, cloud computing comes with a certain risk, for example the aspects mentioned in Figure 2: hardware is owned by and located at the CSP, resources are shared with other customers and data is transported over the public internet. For many organizations, these are reasons why they do not want to use the cloud for (all of) their IT services. A 76% majority of participants in the KPMG cloud computing survey [5] considers security issues to be their main concern regarding the use of cloud computing. In addition, the participants consider legal (51%), privacy (50%) and compliance issues (50%) to be areas of risk.

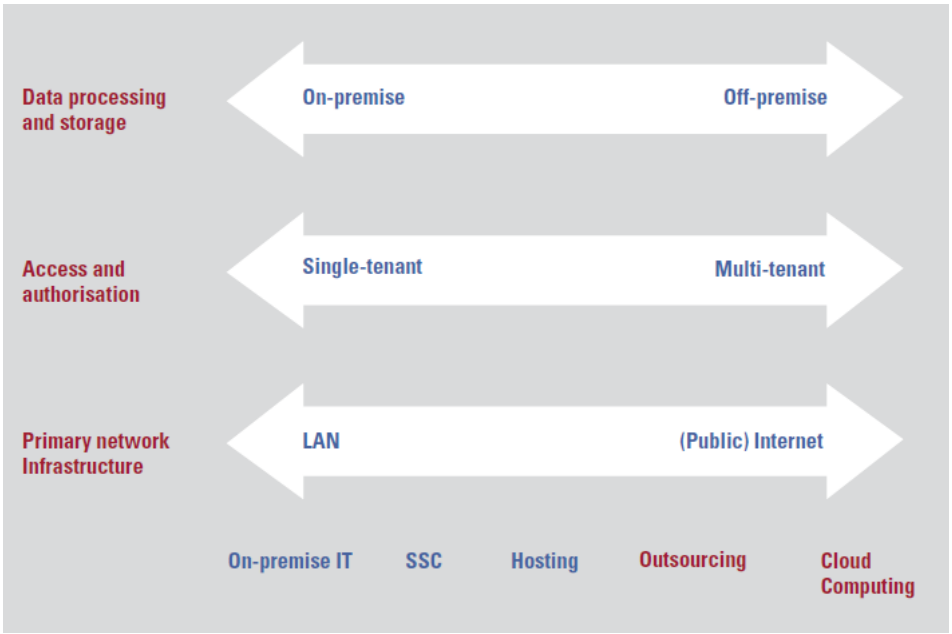


Figure 2 Cloud computing aspects [2]

Interviews with KPMG experts on the cloud market show that private consumers embrace cloud services. However, businesses users are not adopting cloud services; before organizations can move

to the cloud, a number of requirements has to be met. One of the requirements is that the organizations still conform to all applicable regulations and legislation. An important aspect that hinders businesses users from going to the cloud is compliance to data location legislation [5].

1.1.3 Data location legislation

EU Directive 95/46/EC [9], better known as the EU Data Protection Directive, is part of the European privacy legislation and regulates the processing of personal data within the European Union. Personal data is defined as "any information relating to an identified or identifiable natural person". The EU directive makes a difference between the "controller" of the data, who determines the purposes and means of the processing of personal data (data owner), and the "processor" of the data, who actually processes and stores the data. The responsibility for compliance rests on the shoulders of the controller. The data protection rules are applicable not only when the controller is established within the EU, but whenever the controller uses equipment situated within the EU in order to process data. Controllers from outside the EU, processing data in the EU, do also have to comply to data protection regulation.

In an example where a Dutch car dealer stores information about his clients in a database that is managed by an external IT company, the car dealer is the controller and the external IT company is the processor. In this example, the car dealer has to show compliance to the EU Data Protection Directive.

Chapter IV, article 25-26 of the EU Data Protection Directive [9] states that personal data may only be transferred to a country outside the European Union, if that country provides an adequate level of protection. Only four countries are listed as having an adequate level of protection. The result of this directive is that organizations located in the EU or organizations processing data in the EU, must know the location where their data is stored and processed. See Appendix B for the complete text.

An interview with an Account Executive Public Affairs at Fleishman-Hillard (see Appendix D), shows that the European Commission is currently reviewing the European Data Protection Framework. A hearing held with Viviane Reding (the European Commissioner in charge of the review) in March 2011 shows a number of new and important developments in the review: the new legal data protection framework will apply to all EU citizens regardless of where the data is collected and stored. Translated this means that EU data protection rules will also apply to organizations that process and store data of European citizens but are based outside of the EU.

The Dutch implementation of the EU Data Protection Directive is the "*Wet bescherming persoonsgegevens*" (Wbp) [10]. The processing of personal data should be reported to the "*College Bescherming Persoonsgegevens*" (CBP, formerly known as "*Registratiekamer*"), which stores the registrations in a public register and monitors compliance with Wbp. The Wbp consists of the same content as the EU Data Protection described before. Chapter 12 of the Wbp states that organizations in the Netherlands may not transfer personal data outside the EU.

The United States is not listed as a country with an adequate level of protection. For storage of personal data in the United States, the Safe Harbor Principles were developed [11]. Organizations that can show that they have an adequate level of protection are added to a list that is maintained by the US government. For companies in the EU, it is allowed to store and process personal data at companies on the Safe Harbor List.

Due to legislation, companies have to store privacy sensitive information within the EU, or other countries that provide a certain level of minimal protection. This holds for any type of storage: when data is stored on paper, within own IT systems, and on third party IT systems. The next subsection relates this legislation to cloud computing.

1.1.4 Current situation

In the case of cloud computing, the customer of cloud services is the controller, and the cloud service provider (CSP) is the processor. As mentioned in the previous section, the customer has to be compliant to the EU Data Protection Directive, and has to show that privacy sensitive data stays within the EU.

At the moment, it is difficult for cloud customers to determine what happens with their data that is stored in the cloud, because customers do no longer have (direct) control over physical servers, security measures and data location, so the customer has to trust the CSP. This is especially difficult regarding compliance; when the customer does not know the location of its data, it cannot show compliance to the EU Data Protection Directive.

In IT, it is common to have a service level agreement (SLA)in which the CSP and the customer make agreements on a minimum level of the quality of service and additional arrangements. However, for most of current well-known cloud services, customers can only accept standard, non-customizable SLAs. In these SLAs, CSPs offer certain guarantees like uptime, but other aspects like data location are not mentioned or guaranteed. E.g. Google Apps offers only one standardized SLA for all its customers [12], Salesforce.com does not have a SLA at all [13] and Microsoft Office 365 did not provide a SLA during the beta phase. Office 365 however will provide EU data location guarantees when the product is out of the beta phase.

An example that demonstrates this problem is the Dutch government [14], which has defined a 'cloud first strategy': all government ICT has to be taken from the cloud as much as possible; only with good arguments this rule can be deviated. However, the Dutch government has concluded that the cloud market is not mature enough yet to be able to show compliance to legislation, so it will not use any public cloud service, but it will build its own private government cloud.

1.1.5 Conclusion

2011 is 'the year of cloud computing', customers rapidly adopt cloud services. However, businesses are not using cloud services that much. One reason for this slow adoption is legislation that applies to these businesses. They have to store information conform this legislation, which means in case of e.g. the EU Data protection directive that data should stay within the EU. However, current market offerings do not always comply to this legislation, or in case they do, show this compliance to customers.

1.2 Document structure

The rest of this document is structured as follows. Chapter 2 gives theoretical background information about cloud computing. Chapter 3 gives an overview of the research methodology, by linking the research questions to research methods. Chapter 4 introduces the new demands cloud customers have in a cloud computing environment. Chapter 5 describes the current situation at CSPs and describes the typical technical infrastructure CSPs use. Chapter 6 investigates what the current limitations are for CSPs to show compliance to data location. Chapter 7 describes techniques for

negotiating and enforcing security policies to overcome the limitations. Chapter 8 combines the gathered information into a new Cloud Computing Compliance Guideline, which should help CSPs in showing compliance to customer demands regarding data location. Chapter 9 validates whether CSPs think this model is feasible. Chapter 10 concludes this research by answering the research questions and providing points for future research.

2 Background

The motivation for this research has been explained in the previous chapter. To get a better understanding about cloud computing, this chapter provides background information about the cloud computing service models and cloud computing deployment models. Understanding the different cloud computing models provides more insight in the problems that occur concerning data location compliance.

2.1 What is cloud computing

To formally describe cloud computing, the definition by the National Institute of Standards and Technology (NIST) is often used, and is used in this research:

DEF1: *"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [15]*

The primary idea in cloud computing is that organizations do not longer manager and own their IT, but have it delivered as a service by a CSP. Over the last years, there is a trend to outsource more and more of IT to external parties. It is difficult to make a sharp distinction between shared service centers, hosting, outsourcing and cloud computing. Figure 3 shows the difference between these terms based on three aspects: delivery of service, management of IT resources and ownership of assets. The more these aspects can be plotted to the right on the arrows, the more can be spoken about (public) cloud computing.

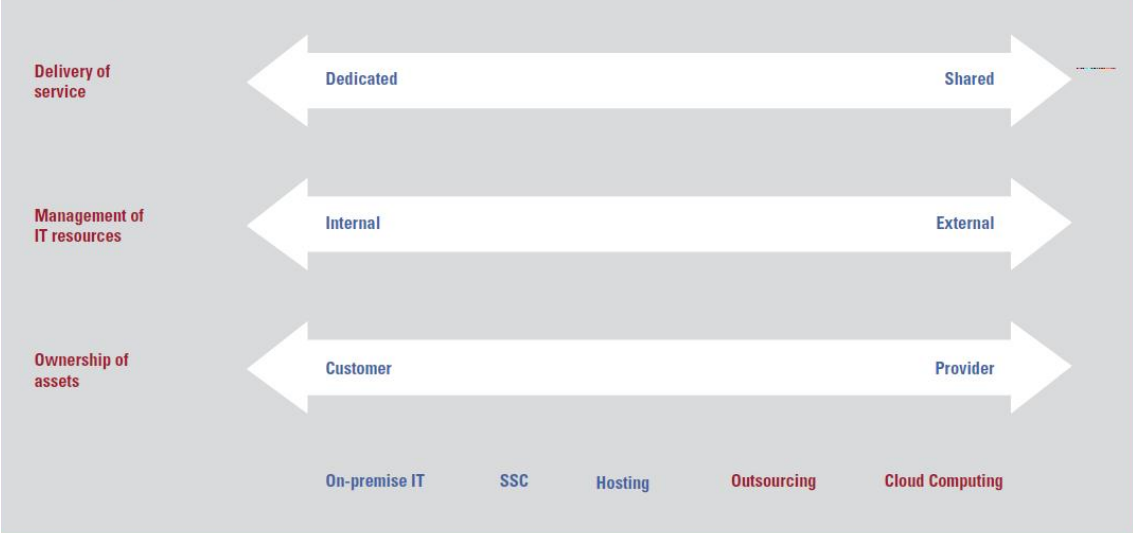


Figure 3 Hosting, outsourcing and cloud computing [2]

To describe cloud computing, and the fundamental difference with traditional IT or outsourcing, the following characteristics [15] can be used:

- **Resource pooling**: contrary to traditional IT, resources are shared by multiple customers (multi-tenancy).

- **Rapid elasticity:** cloud services can be, easily scaled up and down by the demands of the customer. Quickly and temporary scaling up processing power is called 'bursting'.
- **Measured service**: customers only pay for a service they use ('pay-as-you-go' or by subscription) instead of paying for long-term licenses and/ or investments in hardware which are not related to the actual usage.
- **Broad network access:** although leased lines and proprietary networks can be used for cloud computing, its primary infrastructure is the public internet.
- **On-demand self-service**: in contrast to the vast majority of traditional IT, cloud services can be used almost instantly.

An easy to understand example of cloud computing is e-mail. In the traditional IT model, organizations had their own e-mail servers, which were managed by company IT administrators. The e-mail was only available within the office, and the IT administrators had to manage and backup their e-mail for the whole organization. When a server reached its capacity, the administrators had to deploy extra servers. With cloud computing, organizations buy e-mail as a service from a CSP, e.g. Gmail or Microsoft Office 365. The CSP stores the e-mails somewhere on its servers, manages the backups, and delivers a nearly 100% availability from anywhere over the world. And when an e-mailbox is full, it is easy and cheap to buy some extra storage space. The organization only pays for the amount of service it uses.

2.2 Service models

To be able to talk about more specific services, cloud computing can be split into three service models, Software, Platform and Infrastructure as a Service [15]. These service models describe the degree of service / control the CSP offers, and the degree of freedom a customer has. Figure 4 gives a graphical representation of the different service models, and their components. The blue blocks (indicated with 'you manage') are managed by the customer, grey blocks (indicated with 'delivered as a service') are delivered as a service by the CSP.

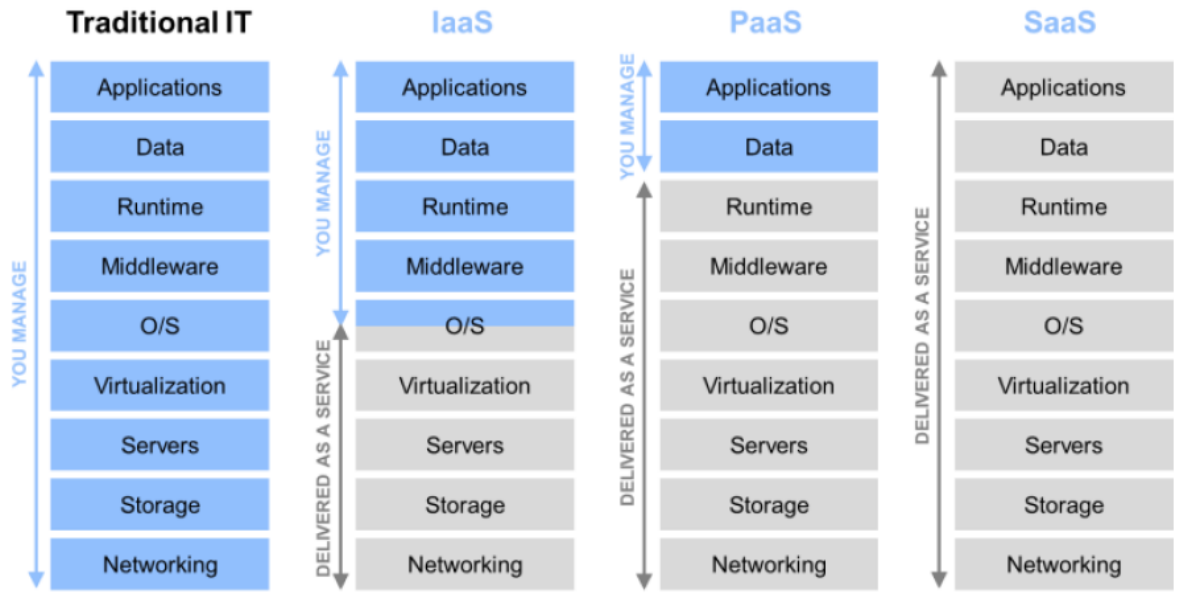


Figure 4 Cloud computing service models [4]

To explain the different service models, a company which uses a Customer Relationship Management (CRM) application is used.

2.2.1 Traditional IT

In the traditional IT environment, all computing infrastructure is located and managed on-premise. An organization buys its own servers, IT administrators manage the complete infrastructure from networking to application levels.

In the CRM example, the company IT department buys servers for the CRM software, installs the operating system, and deploys the CRM application on the server and client computers. Backups are managed by the IT department, and also expansion of the capacity. The company pays for the buying of new servers and licenses for the CRM software.

2.2.2 Infrastructure as a Service (IaaS)

Using Infrastructure as a Service (IaaS), the customers buys infrastructure services from a CSP, but manages the layers on top of the infrastructure itself. "In this service model, the CSP offers processing power, storage, networks, and other fundamental computing resources. The consumer is able to deploy and run operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g. firewalls)." [15]. Examples of IaaS are Amazon Elastic Compute Cloud [16] and Terremark Enterprise Cloud [17].

When a customer uses IaaS in the CRM software example, the customer buys computing power and storage from the CSP. The customer IT department administrators configure a virtual machine on the infrastructure, on which an operating system is installed. They deploy the middleware for communication with other applications, and install the CRM software. There is no need to buy extra servers, when the application needs more resources, extra CPUs and storage can be assigned via a web interface or via the CSP, the customer only pays for the used computing power and data storage.

2.2.3 Platform as a Service (PaaS)

In the Platform as a Service (PaaS) model, the CSP offers a development platform on top of the services delivered with IaaS. "The consumer is able to deploy applications onto the cloud infrastructure created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including, but has control over the deployed applications." [15]. Examples of PaaS platforms are Amazon Elastic Beanstalk [18], Microsoft Azure Platform [19], Force.com [20] and Google App Engine [21].

The CSP offers a development platform, on which applications can be built. This means that the customer IT department has to develop the CRM software in a programming language that suits the CSP development platform. Developers can take full advantage of cloud opportunities like distributed programming and parallel programming for scalable applications. The platform also enables the developer to deploy the application. The company does not own any servers and pays for the used computing power.

2.2.4 Software as a Service (SaaS)

In the Software as a Service (SaaS) model, the CSP offers all infrastructure as a service, including the application. "The applications are accessible from various client devices through a thin client interface such as a web browser. The consumer does not manage or control the underlying cloud infrastructure, but may be able to set limited user-specific application configuration settings." [15]. Examples of common SaaS applications are GMail [22], Office 365 [23] and SalesForce.com [24].

With SaaS, the customer takes the full application service from the CSP. The customer IT department does not have to install or deploy any software, the application can be used via the internet. The customer IT department (or business analysts) can configure the application to the customer's needs, but only within the boundaries offered by the CSP. The customer only pays for the capacity used, this can consist of e.g. the number of users and/or premium options in the software.

2.2.5 'X' as a Service

Many applications can be delivered 'as a service' these days, e.g. business processes, data, identity, etc. [25]. However, these services are not described in the formal definition for cloud computing, as IaaS, PaaS and SaaS cover the majority of services that can be offered by a CSP. Therefore, this research only uses the IaaS, PaaS and SaaS service models.

2.3 Deployment models

Cloud computing can be delivered with four deployment models: private, public, hybrid or community [15]. These deployment models describe who owns, manages and is responsible for the services.

2.3.1 Private cloud

In a private cloud, the services are completely dedicated to the customer, resources are not shared with other customers. "The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise. Resources are dedicated only to the customer." [15].

Figure 5a shows that the private cloud is only used by one customer, resources are not shared with other customers. The cloud service may be offered by the customer's IT department itself, or by an external CSP. The Dutch government is example of an organization which is building its own internal private cloud.

2.3.2 Public cloud

In a private cloud, the delivered services are shared with other customers. "The cloud infrastructure is made available to the general public and is owned by an provider selling cloud services. Resources are shared among all customers." [15].

Figure 5b shows that in the public cloud, resources are shared with multiple customers, which may operate in different market segments, and may have different security demands. Public clouds offer most of the cloud advantages, as the CSP can optimally utilize the resources by sharing them among multiple customers.

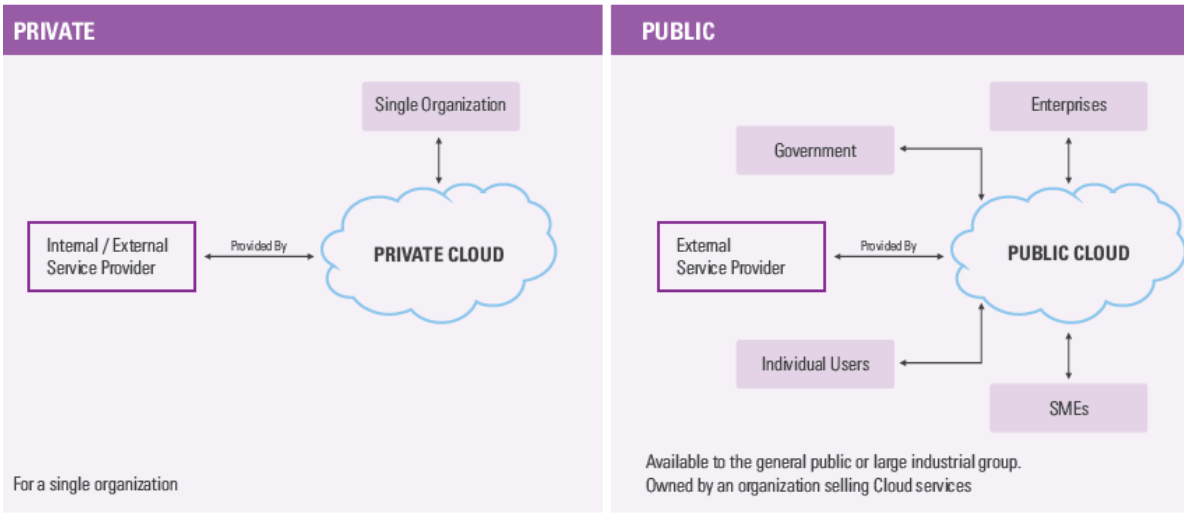


Figure 5 a) Private and b) public cloud deployment models [26]

2.3.3 Community cloud

The community cloud combines aspects of the private cloud and public cloud: resources are shared, but only with other customers that have the same requirements. "The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise." [15].

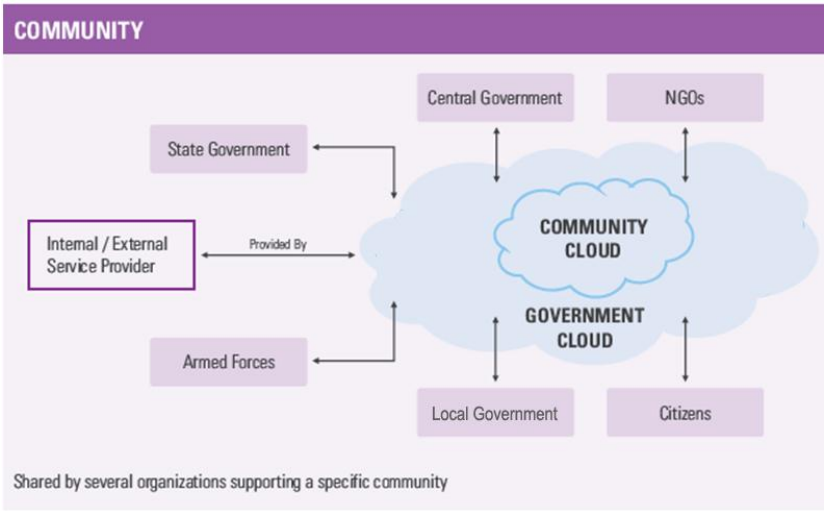


Figure 6 Community cloud computing delivery model (adapted from [26])

Figure 6 shows an example of a community cloud, which is in this case used for a government community. The users of this community cloud (government agencies; all purple blocks in the figure) have the same demands and security requirements for their IT. Google offers such a government cloud with the Google Gov Cloud [27].

2.3.4 Hybrid cloud

A hybrid cloud combines multiple deployment models. "The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds)." [15].

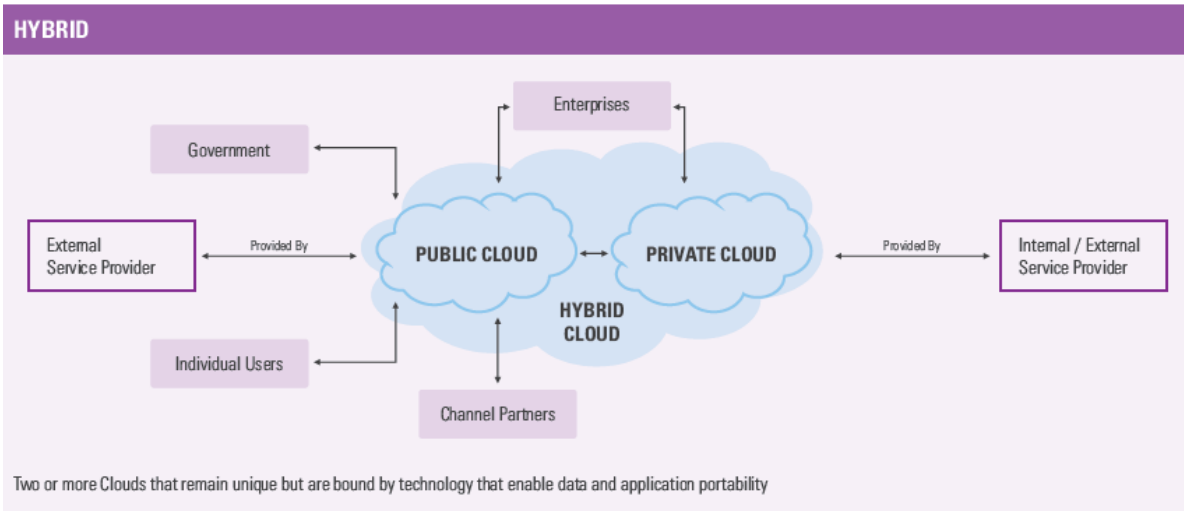


Figure 7 Hybrid cloud computing delivery model [26]

Figure 7 gives a graphical representation of a hybrid cloud, consisting of a public cloud and private cloud. The private cloud is only used by the customer, while the public cloud is shared with other customers. The private cloud and public cloud may be offered by different service providers. At the moment, it is difficult to 'orchestrate' these different clouds, in terms of information exchange and identity and access management [2].

2.4 Conclusion

This chapter discussed the theoretical background and concepts of cloud computing. The combination of service models and delivery models leads to a lot of possible cloud solutions. For data location issues, in particular public clouds are interesting, as resources are shared with multiple customers.

3 Research methodology

This chapter describes the methodology for the research. Section 3.1 describes the scope of the research, section 3.2 describes the problem statement and research questions and section 3.3 describes the methodology used to answer the research questions.

3.1 Scope

This section delimits focus of this research using the following attributes: stakeholder perspective, compliance, customer segment, CSP segment, cloud deployment model and cloud service model.

3.1.1 Compliance aspects

Customers may have several aspects on which they have to show compliance and therefore want guarantees from a CSP on these aspects, like authorization, physical access, data location, employee screening etc. This research focuses solely on the data location aspect, because market experts indicate this as a major barrier for cloud computing [5].

3.1.2 Stakeholder perspective

It is possible to approach data location compliance from a customer perspective, or from a CSP perspective. Customers are the controllers of the data, so they are responsible for showing data location compliance to legislation. However, to be able to do this, customers need information and guarantees from CSPs about the location of their data. The focus of this research is on gathering this information, so the problem is approached from the CSP perspective, as the CSP is the party to give data location information to enable customers to be compliant.

3.1.3 Customer segment

The compliance aspect is mainly applicable to businesses that process or store (privacy) sensitive and confidential data within the EU. This research focuses on enterprises, rather than on individual customers, who do not have to comply to this legislation [9].

3.1.4 CSP segment

This research focuses on CSPs that use market standards for their data centers and software, so the results of this research can be used for all CSPs using the same standards. That also means that CSPs that have developed their own data centers and software, like Google with the Google File System [28], are out of scope.

There is a focus on mid-size CSPs. A focus on small providers would result in providers with probably only data centers in one country. Large-scale CSPs do have multiple data centers all over the world, but they might be reserved in giving away information, and might use their own technology which cannot be reproduced by other CSPs. The compromise is to focus on mid-size CSPs, with multiple data centers, preferable in multiple countries.

3.1.5 Cloud service model

In IaaS, customers can make most of data location decisions their selves. With PaaS, customer have less influence on data location. With SaaS, customers do have the least control over data location, and are most dependent on the CSP. Therefore, the focus of this research is on the SaaS service model. However, the SaaS service model often uses lower infrastructure from the IaaS and PaaS service model, so these service models are also included in some chapters.

3.1.6 Cloud deployment model

The focus of this research is on public clouds, as resources are shared as much as possible there, and customer data is transferred often between resources and possible locations. In private clouds, customer data is processed and stored on resources that are associated with the specific customers, so the data location is clear.

3.2 Problem statement

Before customers can move to the cloud, they have to show that they are compliant to regulations and legislation regarding data location. Customers demand guarantees concerning data location form their CSP, but CSPs do often not offer guarantees about these issues.

3.2.1 Research questions

The goal of this research is threefold. The first goal is to investigate the current situation customers and CSPs experience concerning data location compliance in cloud computing (G1). The second goal is to identify limitations in the current situation (G2). The third goal is to propose solutions for the identified limitations (G3). This research is driven by the following research questions:

RQ1. Which are the typical customers' demands regarding data location compliance?
RQ2. What technical solutions do cloud service providers currently have?
RQ3. What are the current limitations for CSPs to show compliance to customer demands regarding data location?
RQ4. How to make agreements about data location demands between customer and CSP?
RQ5. How can CSPs enforce security policies regarding data location?
RQ6. How can cloud service providers show compliance to customer demands regarding data location in public SaaS cloud computing?

3.3 Methodology

For each research question, a specific research method is used. The research questions and related research methods can be found in Table 1 and are explained in the following paragraphs.

	Question	Method	Chapter
RQ1	What are customer demands regarding data location compliance?	Interviews with cloud experts	4
RQ2	What technical solutions do cloud server providers currently have?	Interviews with CSPs	5
RQ3	What are the current limitations for CSPs to show compliance to customer demands regarding data location?	Literature study	6
RQ4	How to make agreements about data location between customer and CSP?	Literature study, interviews with CSPs	7
RQ5	How to enforce security policies regarding location?	Literature study, interviews with CSPs	7
RQ6	How can cloud service providers show compliance to customer demands regarding data location in public SaaS clouds?	Modeling	8
V1	Validation	Interviews with CSPs	9

Table 1 Research phases, questions and methods

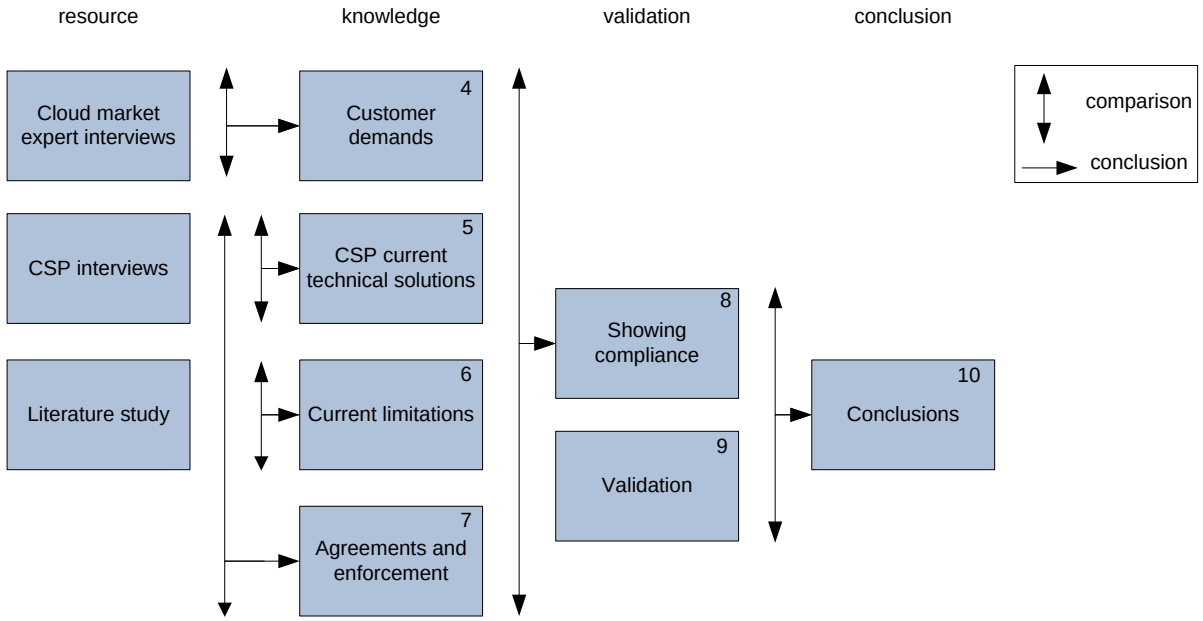


Figure 8 Research model

Figure 8 shows the structure of the outcomes needed in order to reach the goals of this research, according to the technique described by Verschuren and Doorewaard [29]. An arrow in this figure symbolizes a 'confrontation'; a vertical arrow implies items that are compared to each other, a horizontal arrow implies a conclusion. The corresponding chapters in this thesis are shown in the upper right corners of the blocks.

3.3.1 Expert interviews

The first step of this research is to determine the changing demands customers have concerning the new environment cloud that the offers. To gain information about these customer demands, expert interviews are held. It would also have been possible to arrange interviews with actual cloud customers, but because of time constraints, KPMG experts in the cloud market are used.

The interviews have the goal to get an overview of the demands customers have in cloud computing and what the implications are for data location compliance. To achieve that goal, knowledge has to be gathered about how customers determine their demands for cloud computing, how these demands differ from traditional IT, why customers have these demands, and how customers expect think they to be fulfilled by CSPs. The questions for these semi-structured interviews can be found in Appendix E. The results of these interviews are used to determine the current limitations customers experience concerning their demands.

3.3.2 CSP interviews

When information about customer demands is gathered and background information about the agreements process is known, the research focuses on the CSP. To be able to answer RQ2, information is needed about the current situation of CSPs. A first round of interviews with CSPs is held to understand the current situation at CSPs. First, the relation with the clients is discussed, to understand how CSPs experience customer demands, whether customers and CSP see data location as an issue and how the CSP and customer reach agreements on the delivered service. Second, the technical infrastructure is discussed, to understand current limitations in this infrastructure and to

see possible future implementations to show compliance, with a focus on the data location. Third, data location is discussed, to see whether CSPs currently offer services, how they configure their infrastructure and how these security measures are enforced.

The interviews are held in a semi-structured way; a same list with (open) questions is used for each CSP, which can be found in Appendix F. During the interviews, new questions were added to enable a dynamic conversation, which helps to get more in-depth information, when possible.

3.3.3 Literature study

Based on the information about customer demands, and current offerings by the CSPs, a limitations analysis is carried out. The goal is to define limitations CSPs encounter to show compliance to customer demands regarding data location. The limitations are derived from the information gathered during the cloud expert and CSP interviews, and complemented with a search of literature.

To overcome the found limitations, a literature study is carried out. The literature study is carried out using a search on the internet, using related search terms on Google search, Google Scholar and SciVerse. The literature study provides pointers for the implementation of the guideline. Interesting publications are used for a backward and forward scan, to determine other interesting publications.

The goal of the literature study is to answer RQ4 and RQ5:

RQ4: How to make agreements about data location between customer and CSP?
RQ5: How to enforce security policies regarding location?

To answer RQ4, the literature study focuses on security policies to specify security measures, automated negotiation and SLAs. To answer RQ5, the literature study focuses on enforcement of agreements and the enforcement of policies. In addition, the literature study focuses on how to give assurance to verify whether the security policies are actually enforced. The following keywords are used during the literature study:

- (Policy OR Policies) AND (Cloud computing OR Grid computing)
- (Policy OR Policies) AND Specification language
- Service Level Agreements AND Negotiation
- (Enforcing OR Enforcement) AND Agreements
- (Compliance OR Assurance OR Audit) AND Location

3.3.4 Modeling

The result of expert interviews, literature study and CSP interviews are used to define a guideline that describes how CSPs can show compliance to customer demands. The guideline proposes a process for CSPs to demonstrate how to make agreements about customer demands and show compliance to these demands. This guideline focuses on compliance to data location demands, but this may be easily extendable to other customer demands. The guideline also shows which information is needed from a CSP to be able to show compliance to customer demands.

3.3.5 Validation

The proposed guideline is validated to check whether it solves the problem, whether CSPs are convinced it helps them showing compliance to data location and whether it is feasible for implementation in practice. There are two ways of validation: internal validation and external

validation. Internal validation shows that the solution actually works, external validation shows whether the solution still works when the environment changes [30].

3.3.5.1 Internal validation

Wieringa states that "A solution theory is internally valid if its engineering argument is valid when 1) it is true that the interaction among Solution elements and Domain elements will produce certain Outcomes and 2) it is true that these Outcomes will take stakeholders closer to their Goals" [30].

A second round of interviews with CSPs is used to verify the internal validity of the guideline. It is checked whether the guideline is a useful addition to the current situation, and whether it is a feasible to implement the guideline. The results of the interviews are used to improve the guideline.

3.3.5.2 External validation

According to Wieringa, a solution is externally valid "if it is still internally valid when the problem changes a bit. This can be checked with a sensitivity analysis by placing the solution in future scenarios." [30]. During the second round of interviews with CSPs, a number of possible future scenarios with changes in the environment (customer demands, CSP technical infrastructure etc.) is discussed to check the external validity of the solution.

3.4 Conclusion

This chapter discussed the research approach and research methods. This research is driven by six research questions. Cloud market expert interviews are used to get knowledge about customer demands and interviews with CSPs are used to get an overview of the current solutions CSPs offer. A literature study is used to define current limitations, and possible solution theories. The gathered information is used to model a guideline that helps CSPs to show compliance to customer demands. This guideline is validated using interviews with CSPs.

4 Customer demands

As indicated in chapter 1, many organizations would like to move to the cloud, but have concerns about security and compliance. In this chapter, the demands from cloud customers are investigated. This chapter answers RQ1:

RQ1: *What are customer demands regarding data location compliance?*

Section 4.1 describes which customer demands have that are specific to cloud computing, based on cloud markets expert interviews. Section 4.2 gives an introduction to compliance and related legislation and the impact for the customer demands. Section 4.3 describes how customers determine their demands in cloud computing by describing the typical process a customer carries out in before migrating to the cloud.

4.1 What makes cloud computing different for customer demands?

To determine what specific customer demands for cloud computing are, expert interviews are held. Four KPMG experts on the cloud market were interviewed using the interview questions which can be found in Appendix E.

During the interviews, KPMG experts on the cloud market indicated that the migration to the cloud creates extra points of attention for customers, compared to migrations to hosting or outsourcing. The experts indicate that customers attach importance to the following points:

- **Compliance to laws and regulations.** Customers have to comply to applicable laws and regulations. When services are outsourced to a CSP, the customer is still responsible to show compliance, and expects information from the CSP to be able to do that. This point is discussed more elaborately in section 4.2.
- **Data location knowledge**. To be able to comply to legislation, customers need to know the location of their data. The EU Data Protection Directive [9] states that data should be processed and stored within the EU. In addition, customers do not want their data to be stored in countries that have a legislation which allows the government to gain insight into their data, e.g. using the USA Patriot Act [31].
- **Security certificates**. Customers expect the CSP to have an adequate level of security. CSPs can show this using e.g. a SAS 70 certification or ISO 27001 certification. This should also hold for third parties which deliver services to the CSP.
- **Track record of a CSP**. When data storage, storage and management are moved to a CSP, this creates a large dependency of the customer on this external party. Customers demand evidence that a CSP is capable and reliable. An example is data ownership: when the goes out of business or bankrupt, the customer may lose his data or is not able to process the data anymore. The CSP and customer have to make agreements about what will happen in these situation, e.g. by

performing an escrow [1]. Another example is confidentiality of the data; customer require the CSP to protect the data. A CSP can indicate that it handles data secure with security certificates.

- **Cloud readiness of applications**. Not all applications are ready to be migrated to a cloud computing platform. This especially holds for legacy applications, which cannot (or only with large investments) be migrated to the cloud. Customers need to assess which applications can be moved to the cloud, and expect CSPs to guide this process.
- **Internet connections**. Because nearly all cloud computing services are delivered over the internet, the connections between the CSP and customer should be reliable and redundant.

These demands may be different for different types of customers. For example, banks do have strict security policies because trust is an important selling point in the financial sector, while for the shop-next-door these policies are less strict. The level and importance of the mentioned demand depends on the organization.

The KPMG survey on cloud computing shows that compliance and location issues are the biggest barriers for customers to adopt cloud computing; CSPs currently do not offer guarantees on data location compliance. In addition, the other mentioned issues can be solved with currently existing techniques, like escrows or certification for CSP track record and data ownership issues, redundant internet connections for availability and migration and legacy processes for existing applications. The focus of this research will therefore be on data location, compliance and legislation, as there are still research gaps on this topic.

4.2 Compliance in cloud computing

This section discusses what compliance is, which legislation is relevant concerning compliance in cloud computing, what hat the impact is for customer demands and how achieving compliance can be approached.

4.2.1 What is compliance?

The previous section indicated that compliance to legislation is important for customers when considering cloud computing. Compliance is an important term in this research, but this is not strictly defined with a general accepted definition. Compliance is a term that originates mainly in the financial sector, and legislation for financial institutes.

Today, the term compliance is more and more used outside the financial world, with broader definitions. The current general definition is as follows: “Compliance involves ensuring not only that an organization meets the requirements of regulations, legislation, and standards defined by agencies that are external to the organization, but that is also enforces and ensures adherence to its own policies, procedures, standards, best practices, and plans” [32]. In this thesis, mainly compliance to legislation is discussed, as it applies to all customers.

There may be some confusion about the difference between the terms around the concepts of compliance and compliance and security. The following terminology is used during this research:

[1] An escrow is an contractual arrangement made between the customer and CSP, whereby an independent trusted third party receives the e.g. the source code of software. In case the CSP cannot deliver the services anymore, the customer can receive the source code software, so it can keep using the software.

- Customers have to **show compliance** to legislation, e.g. to the EU Data Protection Directive.
- Customers and CSPs make **agreements**, e.g. the CSP will store the customer's data within the EU.
- CSPs have *general* **security policies**, e.g. data center authorization policy, ISO 270001 policies.
- CSPs take *specific* **security measures** for each customer according to agreements between the parties, e.g. configure an environment for specific customer needs.
- CSPs **enforce** these *security measures*, to ensure that the environment is setup conform the agreements with the customer.
- A third party gives **assurance** that the security measures are enforced correctly according to the agreements.
- CSPs **show compliance** to customer demands by allowing a third party audit to give assurance.

Note the difference between customers showing compliance to legislation, and CSPs showing compliance to customer demands.

4.2.2 Relevant legislation

In cloud computing, a number of laws and regulations is important for the customer concerning compliance:

- **EU Directive 95/46/EC (EU Data Protection Directive)** [9]**.** This directive applies to companies which process privacy sensitive data within the borders of the European Union. See section 1.1.3.
- **Sarbanes-Oxley Act (SOx)** [33]**.** The US legislation was enacted as a reaction to a number of major corporate and accounting scandals. It requires companies to manage their IT in such a way that software produces correct financial reports, and changes in software are logged.
- **Health Insurance Portability and Accountability Act (HIPAA)** [34]**.** This US legislation has recently been expanded to include privacy clauses and security requirements for healthcare and insurance organizations.
- **Federal Information Security Management Act (FISMA)** [35]**.** FISMA was introduced in response to concerns about cyber-security. The act requires all federal agencies to develop and implement agency-wide programs to secure data and information systems.
- **Payment Card Industry Data Security Standard (PCI DSS)** [36]**.** PCI DDS is an information security standard for organizations that handle cardholder information for debit and credit cards. The standard was created to increase controls to reduce credit card fraud. Validation of compliance is done annually, by an external assessor for organizations handling large volumes of transactions, or by a Self-Assessment Questionnaire for companies handling smaller volumes.

Some regulations do not specifically regulate the physical location of stored data, although an organization's compliance and security planning may restrict location as part of its strategy. Risk management and data security analysis may be based on the properties of a particular data center. Moving data to a new location may change these analyses, leaving customers non-compliant.

Some legislation and regulations are not directly applicable to European customers, but some of these customers do also need to comply to e.g. United States legislation when they are listed on a US stock exchange.

4.2.3 Consequences of non-compliance

In principle, an organization that stores or processes client data is responsible to show compliance to legislation (in this case the customer of the CSP is responsible to show compliance to its clients). For privacy in the Netherlands, it is the task of the 'College bescherming persoonsgegevens' (CBP) to monitor whether organizations are compliant to privacy legislation. A CBP compliance manual [37] describes three actions that may be carried out when a company does not comply to legislation:

- A citizen can initiate actions
- The public prosecutor may prosecute the company
- The CBP may take legal actions

In all of these cases, when non-compliance has been proven, a judge or the CBP may impose a fine. This shows the need for organizations that handle client data to be compliant.

4.2.4 Legal and regulatory versus accountability approach

Pearson and Charlesworth [38] describe two approaches to accomplish privacy for the customer in cloud computing: the 'legal and regulatory' approach, and the 'accountability' approach. The approach differs per country or jurisdiction, and has consequences for the way to show compliance. The EU Data Protection Directive is an example of the legal and regulatory approach, while accountability is included in privacy legislation in e.g. Canada and the USA, and Pacific countries united in APEC [38].

With the legal and regulatory approach, data location is crucial to enforcement, because the location of data determines the jurisdiction and legislation that applies. With accountability, regulators enforce the law on the 'first in the chain', who has to give the assurance. In this case, data location is less relevant for the customer because of the assurance that data will be treated as agreed regardless of jurisdiction. Because this research often refers to the EU Data Protection Directive, the focus of this research is on the legal and regulatory approach.

4.2.5 Defining location

This section discusses different options how data location can be defined, and how it should be defined in the context of this research. To show compliance to data location, the definition of data location plays an important role. There are various ways to define the location of data. It can be described as: [39]

- a hard disk,
- a SAN,
- a data center,
- a group of data centers,
- a country,
- a geographical region,
- a juridical domain.

The previous section indicated that customers demand to know the location of their data to be able show compliance to legislation. In the mentioned legislation, 'location' refers to a country, or corporation of countries (e.g. EU). This means that for showing compliance, customers do not need

to know the exact location of their data on a specific hard disk, SAN or server, but the country is specific enough. In this research, countries are used to define data locations.

ISO 3166-1 [40] standardizes all countries in the world, and can be by customers and CSPs as a language to exchange countries. A special note is made for the European Union, although it is not officially a country, it is often requested to standardize, so 'EU' is (not officially) reserved for the European Union. ISO 3166-2 describes per country the different states. This is especially useful to define data location within large countries with different jurisdictions like the United States.

4.3 How do customers determine their demands in cloud computing?

Another aspect that needs to be considered for a customer when moving to cloud computing is how to determine his demands. Before customers store data in the cloud, a process is followed to ensure it the data will be stored correctly and compliant to the relevant rules and legislation. Expert interviews showed that the process consists of risk analysis, data classification and taking security measures. When data is stored off-premise, these security measures are negotiated with the CSP, and service level agreements are made to ensure the correct security levels. This section describes this process in more detail.

4.3.1 Risk analysis

Customers use risk analysis to determine how important their data is, how the data should be handled, to whom it may be disclosed, and which security measures the should demand from the CSP. In the case of cloud computing, customers typically carry out a risk assessment before data is moved to the cloud. In this research, it is assumed that this process results in a CIA-classification of data items.

NIST defines risk as: "a function of the likelihood of a given threat-source's exercising a particular potential vulnerability, and the resulting impact of that adverse event on the organization" [41]. To determine the likelihood of a future negative event, it is analyzed how often threats to an IT system together with the potential vulnerabilities will occur. Impact refers to the magnitude of harm to the target that could be caused by a exploiting vulnerabilities. Risk is measured by the product of likelihood times impact; Risk = Likelihood * Impact [41].

Risk analysis can be done in a qualitative way and a quantitative way. In a qualitative risk analysis, an estimation of the impact and likelihood of the risk is made, e.g. in terms of a scale of high, medium and low. In a quantitative risk analysis, the impact and likelihood are quantified in measurable criteria, usually calculated using financial consequences.

The risk analysis results in a set of risk indicators, which show whether data is crucial to the organization, and the impact of negative events. Risk indicators can also include consequences of non-compliance. With the gathered risk indicators about possible threats, the data can be classified, which is described in the following section.

4.3.2 Data classification

To be able to determine correct security measures for different types of data, data needs to be given a classification. Data within the same class need to have the same level of security, and will be treated with the same security measures.

In this research, it is assumed that the CIA quality aspects are used for the data classification. Other data classification techniques are given by e.g. the Dutch "College Berscherming Persoonsgegevens" [42]. Because the CIA quality aspects are widely used and also is used within KPMG, these aspects are used for classification in this research project.

The NIST 800-60 guideline [43] gives a guideline for security categorization of information and information systems, based on two US federal information standards: the Federal Information Security Management Act [35] and the Federal Information Processing Standard [44]. It states three security objectives (CIA):

- **Confidentiality:** Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information.
- **Integrity:** Guarding against improper information modification or destruction, and includes ensuring information non-repudiation and authenticity.
- **Availability:** Ensuring timely and reliable access to and use of information.

Based on the risk indicators about possible threats determined during the risk analysis, all data is given a rating for each of these three security objective, ranging from 1 (low) to 3 (high). A rating is low (1) if the loss of confidentiality, integrity, or availability could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals. A moderate (2) rating is assigned if the loss of confidentiality, integrity, or availability could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals. With a high (3) rating, the loss of confidentiality, integrity, or availability could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals. The combined rating for confidentiality, integrity and availability determines the classification of data [43].

4.3.3 Security demands and Service Level Agreements

Based on the classification of the data, different security measures should be taken for each class of data. For example, in the case of the EU Data Protection Directive, privacy sensitive data should be located within the European Union. In addition, ISO 27001 and ISO 27002 [45] (information security guidelines) can be used to determine which security measures should be taken. E.g. when have a high rating on availability, the network connections for that data should be carried out redundantly. When data has a high rating on confidentiality, it can be encrypted, or stringent access control mechanisms can be used.

In traditional on-premise solutions, these security measures are implemented by the customer itself. For off-premise solutions, which are managed by a CSP, agreements should be made with the CSP to guarantee a minimum level of security that complies to the classification. This is done using Service Level Agreements (SLAs). These SLAs typically describes the three CIA-aspects, but can also contain other agreements.

Which security measures can be taken and which elements should be contained in an SLA is discussed in the next chapter.

4.4 Conclusions

In this chapter, RQ1: "What are customer demands regarding data location compliance?" is answered. Interviews with experts in the cloud market have shown that customers have to show compliance to legislation, customers demand to know the location of their data, customers demand the CSP to have security certificates, customers demand a good track record by the CSP and demand assistance when migrating to the cloud. For this research, the data location compliance aspect is the most relevant.

The EU Data Protection Directive requires customers to store and process their data within the EU. To be able to show compliance to legislation, have to determine which security demands should be requested to a CSP. Therefore, customers carry out risk assessments on data, give data a classification and determine security demands that should be enforced by the CSP. These agreements (e.g. that data is stored within the EU) are formalized in a service level agreements.

5 Cloud Service Provider infrastructure and data location

The previous chapter provided an overview of customer demands in cloud computing: to be able to be compliant, customers may only store data within certain allowed locations. With on-premise solutions, customers can take their own security measures to ensure data is stored in compliant locations, but in cloud computing, the customer depends on the CSP to take these security measures to be compliant. Therefore, the CSP has to actively manage the location of its customers data. This chapter investigates what services CSPs currently have, by giving an answer to RQ2:

RQ2: *What technical solutions do cloud service providers currently have?*

Section 5.1 starts with an introduction to the technical infrastructure of a typical CSP. Section 5.2 discusses how the location of data can be determined using this technical infrastructure.

5.1 Technical infrastructure

This section describes the typical setup of the technical infrastructure of a CSP, based on interviews held with 5 CSPs. As described in section 2.2, CSPs offer different services. Section 5.1.1 describes the main technical driver in cloud computing: virtualization. Section 5.1.2 describes how data is stored in cloud computing. Section 5.1.3 describes how this data storage is combined with a virtualized environment.

Please refer to Figure 4 on page 8 for a description of the cloud service models, while reading the following sections.

5.1.1 Virtualization

The main technical driver in cloud computing is virtualization. This technique allows a CSP to run multiple 'virtual' servers (guests) concurrently on one physical server (host). Such a virtual server is called a Virtual Machine (VM). In the IaaS service model, CSPs offer a VM and all underlying infrastructure, i.e. processing, storage, networking. There are two options in delivering a VM: a client can create its own VM with operating system and configuration, or the CSP delivers a standard VM with pre-installed operating system.

The management of these VMs on a physical server is performed by the hypervisor, also known as virtual machine manager. The hypervisor presents a virtual operating platform to the guest operating systems and manages the execution of the guest operating systems. The hypervisor gives the guests operating systems the impression that they are running on physical hardware, by assigning processing capacity, data storage and networking facilities. Examples of hypervisors are VMware [46], Hypver-V [47] (commercial) and Xen [48] (open source).

5.1.2 Data storage

In data centers, servers do not store data on their own hard disks, but on large storage clusters.

A Storage Area Network (SAN) is a dedicated storage network that provides access to consolidated, block level storage. SANs are primarily used to make storage devices accessible to servers so that the devices appear as locally attached to the operating system. A SAN typically has its own network of storage devices that are generally not accessible through the regular network by regular devices [49].

Note that a SAN alone does not provide the "file" abstraction, but only block-level operations. In contrast to SAN, Network Attached Storage (NAS) uses file-based protocols such as NFS or SMB/CIFS where it is clear that the storage is remote, and computers request a file rather than a disk block.

5.1.3 Data storage virtualization

The hypervisor manages the data storage for the running VMs. The guest operating system thinks that it writes directly to a hard disk, while actually the hypervisor converts these operations to a virtual disk. These virtual disks are often referred to as LUNs[2] (Logical Unit Numbers). A LUN is a logical reference to a portion of a storage subsystem. A LUN can comprise a disk, a section of a disk, a whole disk array, or a section of a disk array in the subsystem. This logical reference, when it is assigned to a server in the SAN, acts as a physical disk drive that the server can read and write to. Using LUNs simplifies the management of storage resources in the SAN.

LUN Type	Multiple disks	Redundant	Improved performance
Simple	✘	✘	✘
Spanned	✓	✘	✘
Striped	✓	✘	✓
Mirrored	✓	✓	✘
Striped with Parity (RAID-5)	✓	✓	✓

Table 2 LUN types and properties [50]

Table 2 gives an overview of different types of LUNs and their properties [50]. The advanced LUN types have the advantage that data is stored redundant, and/or the performance is improved. However, using multiple disks may result in different data locations and non-compliance to location demands, but interviews with CSPs indicated that VMs and attached LUNs always stay within the same data center for performance reasons.

There are two reasons why CSPs would move data to different countries with different jurisdictions. Firstly, dynamically spreading data over multiple locations leads to more redundant and delivers higher availability. When one data center becomes unavailable, other data centers can take over the tasks. Secondly, storing and processing data at different locations leads to more efficiency, when data can be stored or processed at a location with spare capacity or low processing (e.g. electricity) costs for specific moments, e.g. when solar power is available in overcapacity.

5.2 Data location determination

Appendix C shows the typical setup of a specific CSP that uses multiple connected data centers. For performance, efficiency or redundancy reasons, data may travel between these data centers. It is possible that these data centers are located in different countries. In this case it is not evident in which jurisdiction the customer data is located. The determination of the location of data differs per type of CSP; SaaS providers have to take a different approach than IaaS providers. In this section, an

[2] LUN is actually not a correct reference for a virtual disk, because it refers to the addressing of the virtual disk, instead of the virtual disk itself. [49]

overview of the different service models is given to demonstrate the different ways to determine the location of data.

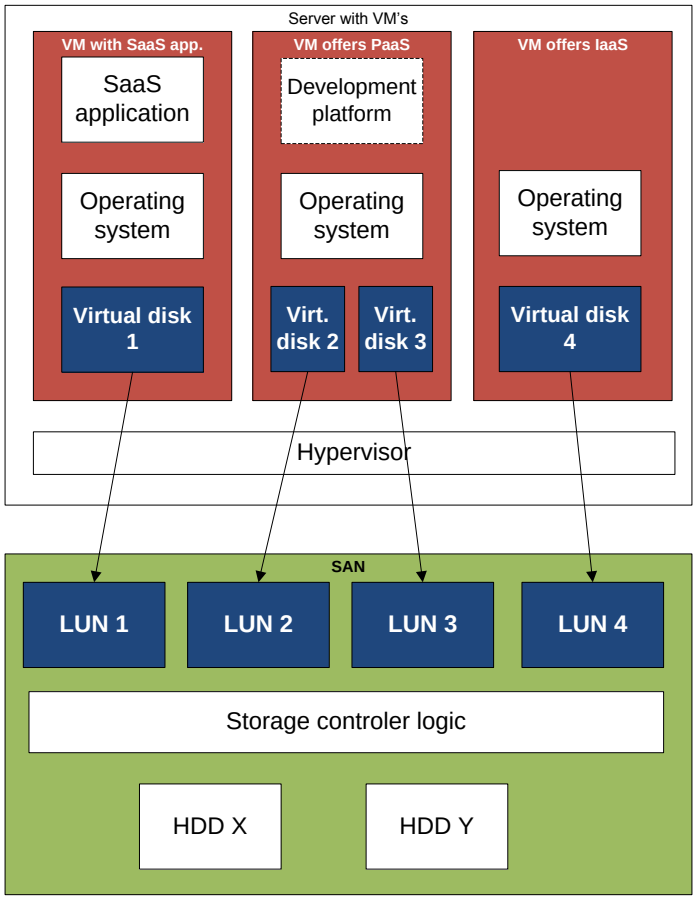


Figure 9 Virtual disks (LUN) in cloud computing (adapted from [51])

5.2.1 IaaS

In the IaaS service model, the CSP delivers computing capacity and storage. Customers deploy their Virtual Machines (VM) with operating system and applications to the IaaS infrastructure. The virtual storage in the VM is attached to physical storage via the hypervisor. Figure 9 shows a VM with a connected SAN for data storage. The hypervisor lets the guest operating system in the VM think it is attached to a physical hard disk (HDD) directly, but redirects these hard disk operations to the connected virtual disk (LUN). This LUN is stored on one (or more) of the physical hard disks in the SAN.

VMware software is software which is often used as a hypervisor at IaaS CSPs. VMware Distributed Resource Scheduler [46] allows creating 'affinity' rules that govern the allocation of virtual machines and storage to physical servers. For example, certain virtual machines can always run on the same server for performance or compliance reasons, or data can be always stored on a specific virtual disk. Alternatively, specified virtual machines can always run on different servers for increased availability.

5.2.2 PaaS

In the PaaS service model, the development platform, computing power and data storage are provided by the CSP. The well-known PaaS providers offer different options concerning location. The

Microsoft Azure platform [52] gives users the choice between three zones: US, Europe and Asia. For the Google App Engine [21], no data location choices are available, Google stores the data somewhere on its worldwide servers. On Force.com [20], users do not have any choice about the data location.

A general problem in controlling the location of the data with PaaS services is that PaaS providers might depend on third parties for their infrastructure (IaaS). This makes it more difficult to detect the location of the data. In case IaaS is used, data location can be delegated to the IaaS provider.

5.2.3 SaaS

The setup of SaaS environments might differ per provider, and with the enormous number of different SaaS applications it is difficult to make statements about data location determination that can be generalized. Some providers offer the same data location availability zones as mentioned with some PaaS providers. SaaS providers may depend on PaaS providers and/or IaaS providers. It is the question whether the data location at these service levels can be controlled by the CSP or the supplier.

A general problem with SaaS providers is that they may use third parties to provide the platform and/or infrastructure. These third parties may also depend on other parties to provider their services. For customers, it is difficult get an overview of this 'chain of suppliers', which makes it also difficult to exactly determine the location of customers' data [53]. However, it stays the responsibility of the CSP to show compliance to customer demands, even when the CSP uses external parties for data storage. This implies that the CSP has to make agreements with its suppliers that conform to the agreements made with the customer.

5.2.4 From virtual locations to physical locations

For all service models it holds that when it is known on which server or data storage cluster the data is stored, it may be still unknown where the data is located geographically. When a CSP has multiple data centers, it is located within one of the data centers, but may be unknown which. Therefore, the CSP should keep track of the geographic location of their assets. This is often done using a Configuration Management Database.

The Information Technology Infrastructure Library (ITIL) [54] is an IT management framework that provides practices for IT services management, IT development and IT operations. ITIL gives detailed descriptions of a number of important IT practices and provides checklists, tasks and procedures that IT organizations can tailor to their needs. One of the aspects of IT service management discussed in ITIL is Configuration management.

The goal of ITIL Configuration Management [55] is enable reliable and accurate information about the IT Infrastructure. This is done by registering and keeping track of changes in the IT landscape (Configuration Items). The administration of Configuration Items is carried out using a Configuration Management Database (CMDB).

The combination of the data located on a specific physical machine and the geographic location of that physical machine in the CMDB allow the CSP to determine the exact location of the customers data.

5.2.5 Data location movement

It is important to keep monitoring the location of data. When data is moved, the agreements with the customer may be broken. Often an SLA states that such a security breach has to be reported to the customer.

The difference per cloud computing service model, as indicated in the previous sections, also holds for monitoring the movement of data location. For IaaS, it is possible for CSPs to monitor the logs of the hypervisor for the movement of data or virtual machines. For PaaS and SaaS, this may depend on the specific implementation of the CSP.

5.3 Conclusions

This chapter answers RQ2: “What technical solutions do cloud service providers currently have?”. The technical infrastructure CSPs currently have enables them to determine and manage the location of data. The determination of data location differs per cloud computing service model; for IaaS the data location can be determined and controlled via the hypervisor, a CMDB is used to determine the geographic location of a server. For PaaS and SaaS this depends on the specific implementation by the CSP and underlying technical infrastructure. Data movement can be tracked by using logging and monitoring tools available for the CSP.

So CSPs are actually in principle able to control and monitor data location, but it is not common practice to offer services on data location compliance. Chapter 6 investigates why this is the case.

6 Current limitations for CSPs in showing data location compliance

Chapter 4 introduced the customer demands in cloud computing. The main point in this chapter is that customers need to be compliant to legislation which requires to know the location of stored data. To be able to do that, customers need their CSP to provide that information. Currently, it is difficult for customers to determine the location of their data, in particular when the offered service has a chain of suppliers. Customers expect the CSP to be able to show compliance to the data location legislation. However, it is not common for CSPs to offer such a service yet.

Chapter 5 provided an overview of the technical infrastructure of CSPs, how location of data can be determined and tracked. This chapter describes the current limitations that CSPs have to show compliance to customer demands by giving an answer to RQ3:

RQ3: *What are the current limitations for CSPs to show compliance to customer demands regarding data location?*

Pearson [56] provides three main open issues in cloud computing related to data location in cloud computing. This chapter discusses these open issues, and describes why these issues hinder CSPs from showing compliance to customer demands regarding data location. Section 6.1 describes the limitation of making clear agreements about data location. Section 6.2 describes the limitations of enforcing agreements and giving assurance about the location of data. Section 6.3 describes the limitation of CSPs using subcontractors, which makes showing compliance even more difficult.

6.1 Negotiation and agreements

To be able to show compliance, the customer and CSP negotiate about the agreements to which the CSP will show compliance.

There is a difference for public cloud environments and private cloud environments. In private clouds, the environment is setup specifically for the customer, so it is easier for the customer to negotiate about specific demands, as these do not impact other customers. In public clouds, a standardized service is offered which is shared with other customers. This makes it more difficult to make specific agreements with the CSP, because this requires the CSP to adapt the environment for each user which requires extra management. CSPs often limit the options that are adaptable by each customer.

Negotiation about agreements can happen in multiple ways, it can be done manually in a conversation between the customer and CSP, it can be done automatically via an interactive form on the CSP's website, or it can be done automatically using a negotiation application or web service. During the interviews with CSPs, it was indicated that currently the intake of new customers and negotiation about demands is a manual process. A presales consultant investigates the customer demands, and delivers a service offer to the client. The advantage of a manual process is that the CSP can help the client with determining the exact demands, because the customer often does not know them exactly, and individual demands can be taken into account. The disadvantage of a manual process is that it may take a long time. The advantage of a automated process is that it can be executed within a short time. However, the state-of-the-art in cloud computing has no or limited support for dynamic negotiation of demands between customer and CSP [57].

The negotiations result in a set of agreements between the customer and CSP in which the delivered service is stated. These agreements are often specified in a contract and/or SLA [58]. A SLA may contain quality benchmarks like uptime, response time to issues, resolution time to issues and other Key Performance Indicators (KPIs). The SLA can also contain information about where the customer's data may be stored, how this is determined and how the CSP reports compliance to this agreement. Such a SLA is often described in a natural language.

To be able to enforce a SLA technically, the SLA needs to be translated to specific security measures and machine readable policies. However, the translation of human readable contracts to machine readable policies has proven to be very difficult [38], although there are several examples of how translations into machine readable policies can be done:

- The IBM Sparcle project [59] investigated the translation from natural language based policies into XML code that can be utilized by enforcement engines. The research seemed to be promising, but the project ended in 2008 without an integrated solution.
- The IBM REALM project [60] investigated the translation of high level policy and compliance constraints into machine readable formats. This research also had promising results, but the project ended in 2005 with only a few research papers.
- Breaux and Antón [61] propose a methodology for extracting privacy rules and regulations from natural language text.
- OASIS LegalXML / eContracts [62] is an OASIS standard for creating and managing contract documents and terms.
- The Encore project [63] is a research project that just started on standardization of technical policies, which can be enforced by multiple parties.

However, interviews showed that these examples are not used in practice yet to translate human readable SLAs in cloud computing into machine readable SLAs. The main reason CSPs indicate to keep using their manual intake meetings is that they need to understand and guide the customer before they can make SLAs. The level of agreements that can be made differs per (type of) CSP. Manual negotiations take time, and lead to contracts and SLAs that have to be translated to be machine readable. Automated negotiation and creation or translation of machine readable SLAs is not available yet.

6.2 Enforcing data location

To show compliance to the agreements made, CSPs have to take security measures, which guarantee a correct enforcement of the agreements. Section 6.2.1 describes current limitations in enforcing these agreements. To be able to show compliance, an independent organization has to give assurance whether these security measures conform to the agreements. Section 6.2.2 describes current limitations in giving assurance.

6.2.1 Enforcing data location

When the agreements are made and documented in a SLA, the CSP has to take security measures to enforce these agreements.

A security measure that can be taken is to encrypt the data. It is debatable whether encrypted data is privacy sensitive, and whether the location encrypted data requires compliance to privacy

legislation. To explain this, two scenarios are described: data encryption applied by the customer and data encryption applied by the CSP.

- In case the customer encrypts the data before it is sent to the cloud, and does not share the decryption key, the CSP cannot retrieve the original data. This has the advantage that the data is stored encrypted at the CSP, so hackers or CSP administrators cannot read the original data, while under the custody of the CSP. In addition, data encrypted at the customer site is not considered as 'privacy sensitive data' in EU legislation, so data location issues do actually not apply for data encrypted by at the customer site. However, when agreements about location have been made with the customer, the CSP still has to show compliance to these agreements. Another point is that it is not possible to take full advantage of cloud computing benefits like scalable data processing, because the CSP is not able to process the encrypted data. For processing, the decryption key is needed, but this directly makes the data privacy sensitive again.
- In the case encryption is added by the CSP, the decryption key is also stored at the CSP. The EU Data Protection Directive still applies to this kind of data, so the location of data is still an issue. CSP encryption does not help in showing compliance, because it does not guarantee the location, it just adds some extra security against intrusion.

It can be concluded that both data encryption by the customer and data encryption by the customer do not help in showing compliance to data location legislation.

A technique to enforce security policies is the 'sticky policies' paradigm [64]. In this paradigm, customers attach security policies for individual files. Customers can define policies which describe who is allowed to process the data, when this is allowed and where this is allowed. One of the sticky policy properties can be the location of the processer (CSP). Before a CSP can process, store or read such a file, it has to request permission from trusted authority. The trusted authority checks whether the CSP is allowed to perform an operation, and only gives permission when this is allowed according to the policy. When the customer defines location as a property, the trusted authority needs to determine the location of the CSP, which can be done based on e.g. the IP-address. Within the sticky policies paradigm however, Identifier-based Encryption (IBE) and Trusted Computing Platform Alliance (TCPA) [65] technology is required for policy enforcement. These techniques require the CSP to become a 'trusted platform'. The architecture of a 'trusted platform' is fundamentally different from existing computing platforms in that it must include security hardware (roughly equivalent to a smartcard chip) that acts as the "root of trust" in a platform. This device is called a Trusted Platform Module (TPM). These platforms help to identify the exact location of the CSP, but require an investment in new hardware which is not standardized in current data centers.

Encryption and sticky policies are techniques to enforce security measures, but these techniques do not work in a cloud computing environment because CSPs cannot take advantage of cloud computing advantages. Due to the dynamic nature of cloud computing and virtualization of processing and storage, CSPs currently do not have the tools to enforce security measures for the agreements about data location per customer [66].

6.2.2 Giving assurance

Compliance can be shown by giving assurance that the CSP executes security measures conform to the agreements. It is possible to define penalties in a contract or SLA in the case a CSP turns out to

be not compliant with the customer demands. However, this does not help in *showing* compliance, and is only a reactive measure. A proactive measure for giving assurance is needed.

A straightforward way to give assurance is by carrying out an audit. There are two options in this case: the customer requires an auditor to check whether the CSP conforms to the agreements made with the customer, or the CSPs requests an auditor to check in general whether the CSP conforms to agreements in general to receive a certain certification. In the first case, the customer pays the auditor to audit the specific agreements with that customer. In the second case, the CSP pays the auditor the check the process of making agreements, enforcing security measures that conform to the agreements.

It is common that an external party carries out the audit [67]. An external party has an independent view on the situation, and can give independent assurance whether the CSP complies to the customer demands or not. However, it requires time for an external party to get familiar with the CSP environment, resulting in higher costs. It is also possible that a CSP internal auditor carries out the audit. An advantage is that an internal auditor is already familiar with the environment, but no independent assurance can be given. Probably for this reason, most customer require independent assurance.

Chow et al. [68] note that the legal implications of data and applications being held by a CSP are complex and not well understood by customers. There is a potential lack of control and transparence when a CSP holds the data. It is difficult to carry out an audit at the CSP, because it is questionable whether there is sufficient insight in the operations of CSPs for auditing purposes. Currently, this insight is provided by documentation and audit manuals. Chen and Yoon [69] describe how audits should be carried out in cloud computing environments. Concerning data location, the paper states: “Data location audits should include all the history of data location following its life cycle. Pay special attention to those data located outside legal territory such as in other states or countries.” However, the paper only hints that auditor should request documentation about the location of data, but does not describe which type of documentation this should be, how this documentation should be checked in practice and how this helps an auditor to give assurance.

Security standards and certifications, like SAS 70 [70] and ISO 27001 [45], cover a lot of security topics, ranging from physical security to management responsibility. Though, these standards do not cover data location, so auditors do not pay attention to this topic when carrying out a SAS 70 or ISO 27001 audit. Currently, there are no specific instructions for audits in cloud computing environments, neither are there audit instructions about where information about data location can be obtained, considering the cloud infrastructure.

6.3 Chain of suppliers

In cloud computing, it is not uncommon that CSPs delegate parts of the execution of the service to other service providers [68]. This creates a chain of suppliers which depend on each other, which might affect the location of the data, so all suppliers should take every security measure to comply with the agreements made with the end customer.

Such a chain of suppliers make it more difficult to check where the data is located, and which supplier is responsible and in control of it.

There are three options to check whether the security measures of each CSP meets the agreements made with the end customer:

- Carve-out method [71]: The auditor only checks the compliance with the agreements at the CSP and does not check this at other suppliers. It is possible to add an extra assurance by asking the supplier(s) to provide an audit report to the CSP's auditors, which show that suppliers are also compliant to the agreements. The advantage is that the auditor only has to check the CSP, and does not have to contact the subcontractors. A disadvantage is that only a part of the assurance can be given, and the reports of other auditors need to be trusted. They might not exactly give assurance about the specific agreements with the end customer.
- Inclusive method [71]: The auditor carries out an audit at the CSP, and at all of the suppliers that the CSP uses. In this case, there should be a clear view on the chain of suppliers. The advantage of this method is that full assurance can be given, because the auditor checks the CSP and all its subcontractors. The disadvantage is that the auditor must identify all subcontractors and carry out audits at all of these subcontractors. This becomes an expensive operations, while it is not guaranteed that all subcontractors will allow external audits. It is questionable whether the CSP and/or the customer are willing to pay for such an exhaustive investigation. In addition, it would be impossible for CSPs to change their subcontractors, as they have to be audited to be able to keep showing compliance.
- Continuous auditing [72]. This is a relatively new technique, which started in the financial sector to automate parts of the auditing. With continuous auditing, automated tools are used to continuously check whether the financial data within the company ins handled correctly, by nearly real-time logging and analyzing each transaction to be able to ensure that the all financial data is still correct and consistent. This technique can also be applied to cloud computing and data location compliance. This can be done by e.g. logging each movement of data, and checking whether this movement is still in compliance with the agreements. The advantage is that the auditing process is partly automated, so manual audits can be carry out less frequently. A disadvantage is that there are currently no continuous auditing tools known for (data location compliance in) cloud computing.

The CSP and the auditor have to make agreements on which of these options is chosen, given the mentioned advantages and disadvantages. With one of these options, the auditor can give the customer (partly) assurance about data location compliance within the chain of suppliers.

6.4 Conclusion

This chapter gives an answer to RQ3: "What are the current limitations for CSPs to show compliance to customer demands regarding data location?". To be able to show compliance, the customer and CSP have first to make agreements on the service to be delivered. Based on these agreements, the CSP can arrange security measures which conform to the customer demands. To be sure that these measures really comply to the customer demands, an audit is carried out check these security measures. When this process is carried out correctly, the auditor gives assurance and the CSP can show compliance to the customer demands.

In this process, three open issues were identified by Pearson [56], which after a literature study resulted in the following four limitations for CSPs to show compliance:

LIM1. The level of agreements that can be made differs per (type of) CSP. Manual negotiation takes time, and leads to contracts and SLAs that are not machine readable. Automated negotiation and creation of machine readable SLAs is not available yet.
LIM2. CSPs do not have the tools to enforce the agreements made with the customers to control the location of the data.
LIM3. Currently, there are no standardized audit instructions to give assurance about data location in cloud computing available.
LIM4. The chain of suppliers makes it more difficult to check where customer data is located, which supplier is responsible and in control for this.

Chapter 7 discusses theories that may solve the limitations about LIM1) making agreements; LIM2) enforcing the agreements and LIM4) the chain of suppliers. Chapter 8 combines all this information in a new process guideline for CSPs to show compliance and addresses the LIM3) audit limitations.

7 Agreements and enforcement

Chapter 6 identified a number of limitations CSPs currently have to show compliance to customer demands regarding data location. This chapter describes literature and theories to bridge these gaps, by giving an answer to RQ4 and RQ5:

RQ4: *How to make agreements about data location between customer and CSP?*

RQ5: *How to enforce security policies regarding location?*

To gather more information about making agreements and enforcing policies, a literature study has been carried out. This chapter describes the results of this literature study. Section 7.1 describes the results for negotiation and agreements, section 7.2 describes the results for enforcement of these agreements. Section 7.3 describes how enforcement can be approached when there is a chain of suppliers.

With the literature and theories gathered in this chapter, chapter 8 integrates this knowledge into a process description for CSPs to show compliance to customer demands regarding data location.

7.1 Negotiation and agreements

Chapter 6 identified the following limitation:

LIM1: *Automated negotiation and creation of machine readable SLAs is not available yet.*

To prevent confusion about terminology, it is good to distinguish the concepts that are used in this section. The following concepts are related to making agreements and implementing security measures to enforce these agreements:

- **Customer demands** describe the security demands customers have.
  Example: data for a certain contract (VM, application, …) may only be stored within the EU.
- **Provider offerings** describing the options a CSP offers.
  Example: data can be stored in the Netherlands, the European Union, China or the United States.
- **Agreements** are made between customer and CSP, specified in a SLA.
  Example: data for this contract will only be stored within data centers located in the Netherlands. Compliance to this agreement will be reported monthly
- **CSP Security measures:**
  - **General CSP security measures:** Measures the CSP takes for general data security.
    Example: ISO 27001 certification.
  - **Specific CSP security measures for a specific customer:** Measures the CSP takes for a specific customer environment.
    Example: setting the configuration for a VM to locate the customer data on a SAN in the Netherlands.

As the mentioned limitation indicates, the difficulty is the transition between these concepts: how to match customer demands and provider offerings, to create agreements? How to transform these agreements into security measures and policies? All of the concepts mentioned above can be represented as policies. This delivers two questions: how to specify and describe these policies technically, and how to do the transitions between the different forms of policies.

In this section, a literature study is carried out. Section 7.1.1 discusses a literature study to find policy specification languages for describing the contents of the policies. Section 7.1.2 discusses the results of a literature study about the transition between the mentioned concepts: how to automate and integrate the negotiation process.

7.1.1 Literature study: policy specification languages

A policy specification language is used to specify a policy in an standard way. For this research, a policy specification language is needed that helps customers to document their demands in a formal way, and helps CSPs to document the services that they offer.

A literature study is carried out to investigate which policy specification languages are currently available. The basis for the literature study is an overview made by the W3C [73]. This overview reviews a number of existing policy specification languages. Related to customer demands and security measures, the following policy specification languages are relevant:

- **P3P** [74] is a protocol that allows websites to declare their intended use of information they collect about browsing users. It is designed to give users more control of their personal information when browsing the web.
- **Ponder** [75] is a language for specifying security and management policies for distributed systems. “Ponder can be used to specify security policies with role-based access control, as well as general-purpose management policies.”
- **Ponder2** [76] is a re-design and re-implementation of the Ponder language. It is a complete framework for policy-based management and not just a policy specification language. This version focuses self-management. ”In contrast to the previous version, which was designed for general network and systems management, Ponder2 has been designed as an entirely extensible framework that can be used at different levels of scale from small, embedded devices to complex services and organizations.”
- **Protune** [77] (PROvisional TrUst NEgotiation) is a policy framework meant to support the creation of policies and advanced policy enforcement point. The framework consists of a declarative meta language for driving negotiation agreements, and integrity constraints for monitoring negotiations and disclosure of credentials.
- **Rei** [78] is a policy language based in the OWL Web Ontology Language, that allows policies to be specified as constraints over allowable and obligated actions on resources in the environment. Rei includes meta policy specifications for conflict resolution. The Rei engine reasons over Rei policies to provide decisions whether the behavior of an entity is allowed.
- **SAML** [79] (Security Assertion Markup Language) is an XML-based framework for communicating user authentication and authorization data between security domains. SAML is an XML-based protocol that uses security tokens to pass information about an end-user between an identity provider and a web service.
- **XACML** [80] (eXtensible Access Control Markup Language) is an OASIS standard that describes both a policy language and an access control decision request/response language. The policy language is used to describe general access control requirements. The request/response language lets one form a query to ask whether or not a given action should be allowed, and interpret the result.

To determine which policy specification language is suitable for this research, a number of criteria are applied. The policy specification language is used by customers and CSPs to specify the customer demands and CSP service offerings, the policy specification language should be able to handle customer demands and CSP service offerings. Because this data location is the focus demand and offering in this research, the specification language should be able to contain data location information. When including data location is not possible, the specification language should be extendable with custom parameters, so data location could be included later. In addition to the negotiation and making agreements, the next step in the process is that CSPs enforce the agreements. It would be desired when the policy specification language is prepared to enforcement by the CSP.

These considerations result in the following criteria for the policy specification language:

- Should be able to specify customer demands and CSP service offerings
- Should be able to include data location OR
  - Should be extendable with extra parameters (to include data location)
- Should preferably be prepared for enforcement by the CSP

Table 3 Comparison table of policy specification languages

	Specification of demands and offerings	Data location included	Extendable	Include enforcement
P3P	✘	✘	✘	✘
Ponder	✓	✘	✘	✘
Ponder2	✓	✘	✘	✘
Protune	✓	✘	✘	✘
Rei	✓	✘	✘	✘
SAML	✓	✘	✓	✓
XACML	✓	✘	✓	✓

Table 3 gives an overview of the policy specification languages and criteria. None of these policy specification languages does have the ability to specify data location as a property. This changes the choice to a policy specification language that is easily extendable, so data location can still be added as a property. XACML is – as its name suggests – easily extendable, and the interviews with CSPs showed that XACML is in some cases already used for other purposes. In addition, it is possible to use SAML as an extension for XACML.

As XACML is the only found specification language that can include customer demands and CSP offerings, can be relatively easily extended with data location properties, and is prepared for enforcement by the CSP, XACML is used as the policy specification language for this research.

7.1.1.1 XACML

XACML is an OASIS standard that describes both a policy language and an access control decision request/response language. The policy language is used to describe general access control

requirements. The request/response language lets one form a query to ask whether or not a given action should be allowed, and interpret the result [81]

The typical setup is that someone wants to take some action on a resource. They will make a request to whatever actually protects that resource (like a file system or a web server), which is called a Policy Enforcement Point (PEP). The PEP will form a request based on the requester's attributes, the resource in question, the action, and other information pertaining to the request. The PEP will then send this request to a Policy Decision Point (PDP), which will look at the request and some policy that applies to the request, and come up with an answer about whether access should be granted. That answer is returned to the PEP, which can then allow or deny access to the requester.

The PEP and PDP might both be contained within a single application, or might be distributed across several servers. In addition to providing request/response and policy languages, XACML also provides the other pieces of this relationship, namely finding a policy that applies to a given request and evaluating the request against that policy to come up with an allow or deny answer [81].

7.1.1.1.1 Data location

XACML does not include data location in the language specification. However, XACML can be extended to support new features and parameters for the policy specification language. An example is the GeoXACML extension [82], which is related to location from an access control to spatial data perspective. However, it does not work the other way around; so it is not able to contain properties to describe the allowed locations of data. To be able to support data location as a part of XACML, an extension has to be developed like this is done with GeoXACML. However, the specification language can only specify data location, it cannot enforce it. To be able to enforce data location, an enforcement framework is needed. How this can be done in combination with XACML (and a data location extension) is discussed in section 7.2.3.

7.1.1.1.2 XACML Integration

XACML policy integration [83] is a process that is developed for distributed systems to match and integrate XACML policies. The solution has two key components: a policy similarity process and a policy integration process. The policy similarity process is the process through which policies are compared with respect of the sets of requests they authorize. Given two policies, this policy determines which is most restrictive. On the other hand, the policy integration preferences is an XACML extension by which a party can specify the approach to be taken if their policies have to be integrated with others.

7.1.2 Literature study: SLA negotiation frameworks

In the previous sections, a policy specification language is chosen, which facilitates the specification of the customer demands and provider offerings in negotiations between the CSP and the customer. In addition, the notation of data location has been defined. This section describes how this specification language and data location definition can be incorporated in a framework that facilitates the complete negotiation process.

A search of literature for automated negotiation resulted in four options: WSLA [84], SLA@SOI [85], and the data protection framework [86]. The options are discussed in the following sections, followed by a comparison.

7.1.2.1 WSLA

The Web Service Level Agreement (WSLA) [84] framework is targeted at defining and monitoring SLAs for Web Services. Although WSLA has been designed for a Web Services environment, it is applicable as well to any inter-domain management scenario such as business process and service management or the management of networks, systems and applications in general.

The WSLA framework consists of a flexible and extensible language based on XML Schema and a run-time architecture comprising several SLA monitoring services. WSLA enables service customers and providers to define a variety of SLAs, specify the SLA parameters and the way how they are measured, and relate them to managed resource instrumentations. Upon receipt of an SLA specification, the WSLA monitoring services are automatically configured to enforce the SLA [84].

Patel et al. [58] relate WSLA to a cloud environment. This paper however mentions that negotiation and SLA establishment are out of scope for the research. It focuses on the deployment and monitoring of the SLA.

7.1.2.2 SLA@SOI

SLA@SOI [85] is a consortium of leading Industrial, Academic and Research Institutes from around Europe. The consortium is committed to research, engineer and demonstrate technologies that can embed SLA-aware infrastructures into the service economy. The 38 month project should be concluded by the end of July 2011. The 11 partners from 7 European countries have an available budget of €15.2 Million, with €9.6 Million coming from the European Commission. The project is currently concluding the research and preparing final publications.

The consortium is developing a exhaustive framework that automates the negotiation, implementation and enforcement of SLAs in e.g. cloud computing, and also takes data location into account. The framework should harmonize the perspective of all stakeholders (CSPs and customers), develop standards for SLA specification and negotiation, and give guaranteed quality of service according to the SLAs [85]. However, important current limitations are that the framework is still in a development phase and not usable in practice. It currently only focuses on IaaS, and not yet on PaaS and SaaS.

This section describes the negotiation phase, enforcement phase and data location specific issues, based on articles published after the second year of the research.

7.1.2.2.1 Making agreements

In SLA@SOI, negotiation is the process by which a group of agents come to a mutually acceptable agreement on a contract the required/provided service should satisfy. There can be multiple scenarios: [87]

- 1(customer) – 1 (provider) negotiation
- 1 (customer) – N (providers) negotiation
- M (customer) – 1 (provider) negotiation
- M (customers) – N (providers) negotiation

The SLA@SOI project currently focuses on 1-1 negotiation, but will extend this in later phases.

First, the customer and the CSP have to define the negotiation objectives, a set of parameters over which an agreement must be reached. For the actual negotiation, WS-agreement is used, but the researchers plan to revise this. Data location requirements can be set using the SLA Template (SLAT). The negotiation process results in a SLA that is signed by both the CSP and customer [87]. Further details are not available yet.

7.1.2.2.2 Enforcement

An important part of SLA@SOI framework is the 'Tashi scheduler' [88], which is responsible for accepting client requests through the CM and finding the appropriate physical machine to create the requested virtual machine. When scheduling the resources, the Tashi scheduler takes into account the related SLAs, and makes sure that these are met. When the SLA is violated, or nearly violated, warnings are being logged.

7.1.2.2.3 Data location

In SLA@SOI, the cloud is considered to be geographically diverse, split over multiple data centers, sites and even geographic regions. This brings the added complexity of meeting legal requirement which can vary by jurisdiction. From a physical perspective, for research purposes the cloud is assumed to be divided into logically managed security enclaves.

To do this, metadata is used to describe additional properties of the physical servers including location. This allows the Tashi scheduler to make decisions about the type of services which are allowed to run on them. The location property can be used to refer to the data center, the site, the country or geographic region [39].

However, at the current time, measurement of many of these terms is not yet implemented and is planned for year 3 or even later, as this technology is only beginning to be investigated and may not be feasible to implement during the project.

7.1.2.3 Lin and Squicciarini

Lin and Squicciarini [86] propose a data protection framework that addresses challenges during the life cycle of a cloud service. The framework consists of three elements: policy ranking, policy integration and policy enforcement. These three elements helps customers to find a CSP that can meet its demands, by making agreements about the service delivered and by enforcing these agreements. However, the paper indicates that it mainly describes a vision, and does not focus on detailed techniques of each elements. The following sections describe these three elements.

7.1.2.3.1 Policy ranking

The policy ranking model uses the customer demand policy as input matches this with CSP offerings policy. This policy comparison is made based on a similarity score: when more similarities are found between both policies, the similarity score becomes higher. Such a comparison method is described in [89]. It is common that the customer chooses the CSP that offers the most similar offerings policy. However, both policies do often not match completely, so they have to be integrated to combine demand and offerings.

7.1.2.3.2 Policy integration

It can be assumed that the customer demand policy and CSP offering policy do not match exactly, so an agreement must be reached on the security options. A policy integration module takes all

customer demands and CSP offerings as input, and helps to generate policies to be adopted by both parties. It needs to solve possible conflicts and achieve harmony on all requirements to automatically generate actual policies as output. The Multi-Terminal Binary Decisions Diagram [90] is an example of such a policy integration model.

7.1.2.3.3 Policy enforcement

Once the policies have been created, correct enforcement is demanded to guarantee the protection promised by the policies. The data protection framework describes two approaches for enforcement models, a *tight coupling approach* and a *loose coupling approach*.

With the *tight coupling approach*, the policies are stored with the according data as sticky policies. A solution could be by a combination of Java policies for authentication, to manage data protection using nested JARS, and mapping the access control rights in terms of programmable constrains.

With the *loose coupling approach*, the policies are stored at a separate, centralized location, which makes it easier to update the policies and accessible for the chain of suppliers, while keeping the data portable.

7.1.2.3.4 Combination with XACML specification language

The L&S framework provides a complete framework description for SLA management and integration, but does not have a concrete implementation. Section 7.1.1 showed that XACML is an exhaustive specification language for describing all kinds of policies. In addition, data location is defined in an ISO standard as described in section 4.2.5. When these three concepts are integrated (L&S framework + XACML specification language + ISO language standard), a workable framework may appear. In the rest of this research, this combination is referred to as the 'XACML framework'.

In this XACML framework, the customer demand policy and CSP offerings policy are described in XACML. Following the structure of the L&S framework, these policies are matched to each other, and the best match is used as basis for the SLA. The policy matching and integration can be done using the XACML policy integration technique, as described in section 7.1.1.1.2. The policy enforcement of the SLA can be done using by configuring the hypervisor, which is described in section 7.2.3. The framework does not contain techniques for SLA monitoring or violation detection.

7.1.2.4 Comparison

To determine which SLA negotiation framework is suitable for this research, the frameworks are compared based on a number of criteria. The SLA negotiation framework will be used for negotiation of SLAs. An important aspect of negotiation in this research is data location. After the SLA has been negotiated, the CSP has to enforce the agreement. It is desired when the negotiation framework can already define measures that should be taken to be able to enforce the agreements. Finally, the current status of the framework should be taken into account, is it ready to use by CSPs, or is it still in a academic research phase?

These considerations result in the following criteria for the SLA negotiation framework:

- Is negotiation included?
- Can service level agreement be defined?
- Is enforcement of the agreements included?
- Is data location included as a specific parameter?

- What is the development status of the technology?

Table 4 compares these aspects to the different negotiation frameworks.

Table 4 Comparison table of negotiation frameworks

	WSLA	L&S	SLA@SOI
Negotiation	✘	✓ (policy matching)	✓
Defining SLA	✓	✓ (policy integration)	✓
Enforcement	✘/✓ (monitoring)	✘/✓ (only described in concepts)	✓
Location	✘	✘	✓
Status	Unknown	Still in research	Still in research

WSLA is a framework for describing and monitoring SLAs. However, it does not facilitate the negotiation process or enforcement, it just gives the possibility to describe and monitor SLAs.

The L&S framework does provides a complete framework description for SLA management and integration, but does not have a concrete implementation. However, such a concrete implementation can be achieved using the XACML specification language, XACML policy integration and ISO language standard. When these three concepts are integrated (L&S framework + XACML specification language + ISO language standard), a workable framework may appear. This combination is referred to as the 'XACML framework' for the rest of this research.

On the other hand, such a framework actually already is described, as the mentioned SLA @SOI. This framework consists of all aspects of SLA management, negotiation, defining an SLA and enforcement. Negotiations are based on an interactive process, which requires multiple interactions between the customer and the CSP. However, this project is still in academic development phase, so it cannot be used in practice yet. The project shows promising results, so it seems like a good framework that can manage the complete negotiation and agreements process.

7.1.3 Conclusion

Based on the comparison, SLA@SOI and the XACML framework are the most feasible SLA negotiation frameworks. These two options are considered as solutions that can be used by CSPs for negotiation and making agreements. Both SLA@SOI and the XACML framework achieve the same goals, but have a different approach for negotiation and making agreements. SLA@SOI uses an iterative process that requires frequent interaction between the CSP and the customer for negotiation and making agreements, while the XACML framework uses policy matching and policy integration, which at once chooses the closest matching possible options from the customer demands policy and CSP service offerings policy.

While SLA@SOI and the XACML framework have a completely different approach, they both are a framework that enable CSPs in automated negotiation and making agreements. The XACML framework may be better because its components are already used in practice, while SLA@SOI delivers a better integrated solution, but is still in research.

7.2 Enforcing agreements

Now two SLA negotiation frameworks have been chosen, the next step for the CSP to enforce these agreements. The CSP has to take security measures to do this. Chapter 6 described the following limitation:

LIM2: *CSPs do not have the tools to enforce the agreements made with the customers to control the location of the data.*

This section describes theories how CSPs can enforce agreements. The previous section mentioned two techniques to make agreements: SLA@SOI and the XACML framework. This section describes for both SLA negotiation frameworks how agreements can be enforced.

7.2.1 General enforcing techniques

Most enforcement techniques for cloud computing describe how to enforce customer isolation, so customers are not able to access each other's data on networking and storage level [91]. However, it is difficult to define when data location is enforced correctly. During the research, the following two approaches for enforcing data location were suggested [58] [69]:

- Ensure there are procedures for correct configuration settings, so the hypervisor will allocate client data to allowed locations.
- Enable monitoring and logging of movements of data (location) by the hypervisor. In case the data is relocated to an unwanted location, this could be prevented, or reported as a violation.

In both cases it is important that the CSP exactly knows where the physical hardware is located. In a virtualized environment, it has to be clear on which physical machine a virtualized entity is running.

As discussed in section 5.2.4, the CSP has to keep track of all hardware in the CMDB. The CMDB can be used as input for the enforcement of policies, to check whether a virtualized entity with specific customer data is allowed to be executed on physical hardware with a specific location.

7.2.2 SLA@SOI

To be able to enforce compliance to the signed SLA, SLA@SOI has developed its own infrastructure to manage and deploy Virtual Machines. It first converts the metrics mentioned in the SLA into configurations of its technical environment [92]. Based on this configuration, SLA@SOI uses the developed Tashi scheduler to continually ensure that the configuration is carried out correctly, implying that the SLA is enforced. The configuration includes location, so this automatically incorporates enforcement of data location. The scheduler takes care of tracing and logging of exceptions of the configuration and violations of the SLA for audit purposes.

7.2.3 XACML framework

To be able to enforce compliance in the XACML framework, the framework configures existing technical infrastructure: the configuration of the virtual disks assigned to VM should be set correctly. This is done based on the agreed SLA, which specifies the allowed locations using the ISO 3166-1 standard. The allowed locations as specified in the SLA can be configured in the hypervisor (VMware, Xen, Hyper-V etc.). As described in section 5.2.1, e.g. VMware Distributed Resource Scheduler allows the enabling of such data storage and processing configurations [46]. The hypervisor can also be used to enable logging and monitoring. Currently, it is unknown whether these hypervisors support automatic detection of violation of the agreements.

7.2.4 Conclusion

It is difficult to define when data location agreements are enforced correctly. Two aspects are important to enforce data location agreements: setting the correct configuration combined with monitoring and logging. Both can be applied to the SLA@SOI framework and XACML framework. For SLA@SOI, which developed its own environment for virtualization, this is already integrated using the Tashi scheduler. For the XACML framework, which uses currently existing infrastructure, the hypervisor can be used for configuring data location and enabling monitoring and logging.

7.3 Chain of suppliers

In enforcing the agreements with the customer, chapter 6 identified the following limitation:

LIM4: *The chain of suppliers makes it more difficult to check where customer data is located, which supplier is responsible and in control for this.*

This section describes approaches to this limitation for each of the cloud service models.

7.3.1 Infrastructure as a Service (IaaS)

For the IaaS service model, it is relatively easy to determine and enforce the location of data. However, the PaaS and SaaS service model depend on the other service models and/or infrastructure to determine and control the location of the data. In general, two cases can be distinguished: the CSP manages the infrastructure supporting the PaaS or SaaS service itself, or the CSP uses another supplier for the underlying services.

In the case that the CSP manages the underlying infrastructure for PaaS or SaaS itself, it can itself ensure the correct configuration of the hypervisor (IaaS), and make sure that the offered platform (PaaS) or application (SaaS) is connected to a allowed virtual storage.

In case an external party supplies the underlying infrastructure, the CSP and external party have to make agreements to make sure the agreements conform to the agreements the CSP has with its customers. Of course, the different customers may have different demands. In the example of data location, some customers may have the requirement to store data in the EU, while others demand the data to be stored in the US. In this example, the CSP can choose an external party with data centers in both the EU and in the US, or the CSP chooses two parties who have a data center in the EU and/or the US.

7.3.2 Platform as a Service (PaaS)

The PaaS service models uses the IaaS infrastructure, and delivers a development platform on top of it. This development platform allows developers to build their own applications, which also offers data storage. The development platform should make sure data location is handled according to the agreements with the customer (developer). The connection between the IaaS and development platform is essential.

However, when building an application using the platform, it is up to the developer to ensure data location. The developer has the choice to use the virtual storage provided by the platform, but the developer can also choose to use other storage sources (e.g. via internet connections). This is outside the scope of the compliance by the CSP.

7.3.3 Software as a Service (SaaS)

For SaaS applications, the security measures should be taken on another level. This level depends on the characteristics of the application.

- For *multitenant* applications (e.g. GMail), where multiple customers share one instance of the application the application should facilitate a part of the enforcement of the agreements. One of the interviews with CSPs showed that the application uses different databases to store data of groups of customers with the same security demands.
- For applications that have *individual instance* for each customer (e.g. Microsoft Exchange Online), the security measures should be pushed down the stack to the runtime level, so that each instance is running on a virtual machine that meets the agreements with the specific customer. In the example of Microsoft Exchange Online, the hypervisor hosts the VM with the specific instance of Exchange Online, which should be configured to comply to the agreements with the customer.
- When only a few customer demand options are possible, a *hybrid* construction is possible. In this scenario, each customer demand option has its own application instance, but can serve multiple customers (multitenant). Application instances are shared by customer with the same demands.

7.3.4 Conclusion

The enforcement of agreements differs per service model. IaaS can be handled by the hypervisor, PaaS can be handled by the platform and SaaS can be handled by the application. It is important that the connections between the service models that depend on each other (e.g. SaaS depends on IaaS) are configured correctly by the CSP.

7.4 Conclusion

This chapter answers the research questions RQ4 “How to make agreements about data location between customer and CSP?” and RQ5 “How to enforce security policies regarding location?”. The literature study resulted in two approaches that can be taken in making agreements between the customer and the CSP, and enforcing these agreements.

The first approach is a combination of the XACML specification language extended for data location, together with XACML integration and the Lin and Squicciarini framework. This components are already used in practice, but not as a combination. The second approach is the SLA@SOI project, which is focused on cloud computing and SLAs and covers all aspects of the process, but this research is still in development.

To be able to enforce data location agreements, it is required that the CSP ensures a correct configuration settings for each customer, complemented with monitoring and logging for detection violations. In addition, the mapping of physical hardware to geographical locations using the CMDB is an important aspect to enable enforcement of data location agreements.

The main difference between the two frameworks for negotiation is that SLA@SOI uses an iterative process between the CSP and customer, which requires multiple interactions, while the XACML framework uses policy matching, which requires only one interaction to match the policies and integrate them based on the best match. For enforcement, both approaches differ in the fact that the XACML framework configures the existing infrastructure (e.g. hypervisor), while SLA@SOI replaces the existing infrastructure with its own infrastructure (e.g. the Tashi scheduler).

8 The Cloud Computing Compliance Guideline

Combining all knowledge gathered in the previous chapters, this chapter proposes a new guideline that can give an answer to RQ6:

RQ6: *How can cloud service providers show compliance to customer demands regarding data location in public SaaS cloud computing?*

This chapter introduces a 'Cloud Computing Compliance Guideline' which describes the complete process of showing compliance. This guideline can be used by CSPs to setup or improve their process of showing compliance to customer demands regarding data location. First, a brief overview of the guideline is given. This is followed by a description of the phases in the process in more detail, while relating the limitations found in chapter 6 to the theories found in chapter 7. The guideline describes how the process steps can be implemented practically using two frameworks: the SLA@SOI framework and the XACML framework.

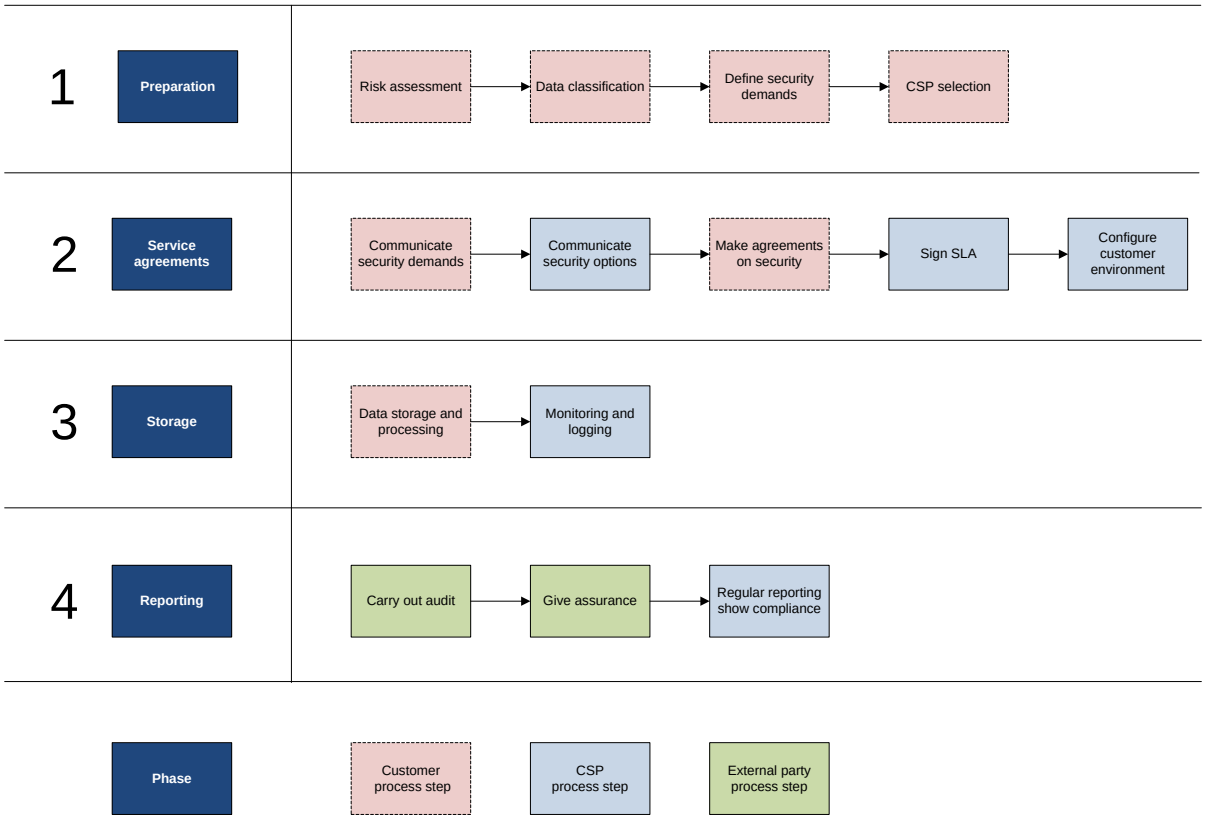


Figure 10 Cloud Computing Compliance Guideline (see also Appendix A)

Figure 10 gives a graphical representation of the Cloud Computing Compliance Guideline. It contains four phases: preparation, service agreements, storage and reporting. The four phases group together a number of process steps. Red process steps are carried out by the customer, blue process steps are carried out by the CSP and green process steps are carried out by an external party (an auditor).

Phase 1 describes how the customer prepares the movement to the cloud, by carrying out a risk assessment, data classification, creating security demands regarding data location and CSP selection.

Phase 2 describes the negotiation process between the customer and CSP. After the automated negotiation, the CSP takes security measures to ensure data will be stored conform the agreements.

Phase 3 describes the regular storage process. Because all security measures are taken, no extra efforts are needed. However, CSP do monitor and log the movement of data, to detect possible violations.

Phase 4 describes how the CSP shows compliance to the customer demands regarding data location. This is done by regularly reporting the current status, and allowing audits to give assurance about the correctness of the process.

8.1 Phase 1: Preparation

When a customer decides to migrate an application to the cloud, a number of steps are taken. First, it has to be determined which data will move to the cloud. A risk analysis of that data is made and the data is classified to determine how the data should be handled concerning confidentiality, integrity and availability. Based on this classification, the customer determines which security demands have to be met. In the case of data location, these demands state where the data may be stored. With this information an overview of CSPs can be made that meet the customer's demands, and the customer has to select a (set of) CSPs. Further details of this process are already described in section 4.3.

8.2 Phase 2: Making service agreements

When a (number of) CSPs is selected, the customer and CSP make agreements on the service level that will be delivered. For this research, in particular data location is important. The CSP has to show which options are available (geographical areas) for which price, and the customer has to indicate the earlier determined demands. Based on this input, the customer and CSP negotiate to reach an agreement.

8.2.1 Negotiation and making service agreements

Chapter 6 described that a current limitation LIM1 is that automated negotiation between customer and CSP and creation of machine readable SLAs is not available yet. Chapter 7 described two techniques for automated negotiation and creation of an SLA: the SLA@SOI framework and the XACML framework. Both frameworks offer a standardized language for negotiation between the customer and CSP. When agreement is reached, SLA@SOI automatically creates a human readable and machine readable SLA. The machine readable SLA can be used to enforce the SLA. For the XACML framework, this process is defined. The CSP should define how these matched policies can be converted into configurations that can be applied to his hypervisor. To define the location of data, the ISO-3166-1 standard can be used.

Off course, it is possible that the customer and the CSP do not reach an agreement. The customer can decide to stop the whole process, or start with (the last steps of) phase 1 again.

8.2.2 Enforcing agreements

To ensure that the SLAs are carried out as agreed on, the SLA needs to be enforced by the CSP. Chapter 6 described limitation LIM2 that CSPs do not have the tools currently to enforce these agreements. Enforcing agreements can be done by ensuring that the configuration is set correctly, and by monitoring and logging whether the SLA is violated.

The SLA@SOI framework uses its own developed environment to control the configuration of virtual machines and uses the Tashi scheduler which automatically takes enforcement into account. The XACML framework configures the excising framework once correctly to enable correct data location constraints and monitoring and logging. This can be done using e.g. VMware DRS as described in section 5.2.1. As mentioned in section 7.3, the approach differs per cloud service model or in case the CSP uses external suppliers to deliver services.

The XACML framework does not solve the chain of suppliers limitation (LIM4) yet, as enforcement is only applied to the CSP's own infrastructure. The SLA@SOI framework however does provide support to delegate tasks to suppliers, but to be able to enforce the agreements, these suppliers also have to use the SLA@SOI framework. Because the SLA@SOI framework uses standardized definitions and communication, it is relatively easy to delegate tasks with the according SLAs to suppliers. The XACML framework does not support this, but can do this when the supplier uses the same infrastructure (and e.g. hypervisor) as the CSP. In that case, the CSP and the supplier can use the same terminology and configuration settings to delegate the tasks with the according SLAs and configuration.

8.3 Phase 3: Data storage

In this phase, the customer stores and retrieves his data in a regular way, without taking extra actions that concern the agreements. Because the CSP has configured the customer environment already in phase 2, the constraints are set so the SLA can in principle not be violated. No extra actions from both the CSP and customer are needed; in the SLA@SOI framework the Tashi scheduler takes care of enforcement, and in the XACML framework the policies are automatically enforced because of the correct configuration of the hypervisor.

However, to continuously ensure enforcement, it is important that the CSP keeps monitoring and logging the movement of the data. In case a configuration is changed manually, or configuration may get corrupt, this should be detected. When such a SLA violation is detected, actions should be taken to solve the violation and inform the customer of this violation.

8.4 Phase 4: Reporting

To show compliance to the customer demands, the CSP should regularly report the current status of the customer's data. This is done with an assurance report containing audit information about the data storage process and enforcement. An auditor regularly checks whether the CSP data storage process conforms to the customer policies.

8.4.1 Giving assurance

As indicated in chapter 6, compliance to the customer demands can be shown by carrying out an audit, which gives assurance about the enforcement of the agreements. In chapter 6, the following limitation for CSPs to show compliance has been identified:

LIM3: *Currently, there are no standardized audit instructions available to give assurance about data location in cloud computing.*

This section describes elements that should be audited to be able to give assurance about the correct enforcement of agreements related to data location. This section only focuses on data

location; other aspects of agreements with customers may also be audited, but are outside the scope of this research.

As described in section 6.2.2, two approaches can be taken: in the first approach a customer requests an auditor to carry out an audit for his own agreements, in the second approach the CSP requests an auditor to carry out an audit for the complete process of enforcing agreements with customers. CSPs indicate that they have a preference for the second approach, because this needs less effort and directly shows compliance for all customers [93].

For the enforcement of agreements with customers, the CSPs often use the Deming circle (plan, do, check, act). In the plan phase, the configuration is prepared, and carried out in the do phase. The monitoring and logging are used for the check phase. In case a violation is detected, the CSP has to take an action to return to a normal state. Auditors have to check whether the steps in this circle are carried out correctly [93]. Note that the auditor does not just carry out the 'check' part of the Deming circle, the complete circle should be executed by the CSP itself; it is the task of the auditor to check whether this circle is executed correctly. CSPs indicate that they also carry out internal audits. This speeds up the process for the external auditor, as the external auditor checks whether the internal auditor carries out the audits correctly, instead of carrying out the audits himself.

For data location agreements, it is particularly important to check the enforcement process. The previous chapter described that configuring, monitoring and logging are important. The auditor should check whether the process of configuring the customer environments is carried out correctly. For the XACML framework, the auditor should check whether the configuration of the hypervisor is correct. For the SLA@SOI framework, the auditor should check the results of the monitoring and logging. In addition, the auditor should check whether the CSP actively logs and monitors the location of data, and detects and reports violations of the agreements. Special attention is paid to the CMDB in section 8.4.1.1.

When the CSP uses third party suppliers to deliver parts of the services (e.g. a SaaS provider may use the infrastructure of an IaaS provider), this relation should be audited by the auditor. This also holds for cases like off-site backups, data replication for disaster recovery and mirroring. As discussed in section 6.3, there are multiple options (Carve-out method, Inclusive method and continuous auditing) to perform these audits. The CSP and auditor make agreements on the type of audits, and report this to the customer.

Besides checking the general Deming process, it is possible to take individual samples for specific customers. To check the complete process for these specific samples, whether they comply to the agreements made, gives an extra layer of assurance [93].

8.4.1.1 CMDB

An audit can be used to verify whether the information in the CMDB still conforms to the actual situation. The general audit guidelines state that audits can take place at the following moments: [55]

- Directly after the implementation of a new CMDB
- Six months after implementation of the CMDB
- Before and after important changes

- After a disaster recovery situation
- On random moments

Though, these are guidelines for audits on a CMDB in general. For data location compliance, it is important that naming of the hardware match the real geographical location. This cannot be done e.g. directly after the implementation of the CMDB, so especially the last three mentioned moments are important to perform an audit on the correct naming of hardware.

Questions that need to be asked during the audit of the CMDB: [55]

- Are all changes in all phases of execution logged in the CMDB, and does Configuration Management controls this?
- Is the actual situation still represented in the CMDB? If not: why is this the case, and what are the consequences for Change Management.
- Does the naming of new Configuration Items still conform to the agreements?
- Are the basic configurations well documented? Can these be used in case of emergency?

For agreements on data location, it is in particular important to check the naming of physical hardware, and relation to geographic areas.

8.4.2 Audit results

When the audit is completed, the auditor delivers a 'Third Party Mededeling' or IT assurance report. There are multiple types of assurance: the report can give a 'reasonable assurance' or 'moderate level of assurance'. The type of report has to be determined before the audit starts. Based on the findings, the auditor can give a: [93]

- Unqualified opinion (Dutch: goedkeurend): the auditor has determined that actual situation does conform to the agreements.
- Qualified opinion (Dutch: goedkeurend met beperking): the auditor has determined that the actual situation does conform to the agreements, but: there was a limitation in scope of investigation, there is remaining uncertainty. When this opinion is issued, the specific reasons for this qualification will be stated.
- Adverse opinion (Dutch: afkeurend): the auditor has determined that the actual situation does not conform to the baseline.
- Disclaimer of opinion (Dutch: oordeelonthouding): the auditor had not enough evidence to be able to give an opinion.

Only with a unqualified opinion and a reasonable level of assurance, compliance to the agreements can be shown. In case of a qualified opinion, it depends on the findings whether compliance can be shown. The CSP should report these findings to the customer.

8.4.3 Iterative loop

When the process ends, the customer has to check whether the shown compliance still meets its demands, as the data confidentiality, requirements or legislation may change. It is therefore advised to regularly restart the complete process again, to do a new risk assessment, and check whether this still meets the current services offered by the CSP. The re-iteration of the process helps both the customer and the CSP to keep their security on a high level.

8.5 Conclusion

To be able to show compliance, assurance about the enforcement of the agreements concerning data location must be given. This assurance can be given by an external auditor, who carries out an audit to check whether the agreements are enforced correctly.

This chapter proposed the Cloud Computing Compliance Guideline, which provides a process description that helps CSPs to setup or improve showing compliance to customer demands regarding data location. The guideline uses the SLA@SOI framework and XACML framework described in chapter 7 to bridge the limitations that were found in chapter 6. The guideline consists of four phases; preparation, negotiation, data storage and reporting. When these phases are carried out correctly, and the auditor gives assurance that the agreements are enforced, compliance can be shown to the customer demands.

The contribution of this guideline is that it describes precisely which steps a CSP should be taken to show compliance to customer demands regarding data location. The guideline can be practically implemented using the SLA@SOI framework or XACML framework. In addition, the guideline describes which points should be taken into account when an audit is carried out, focusing on the enforcement of agreements and connection with the CMDB.

9 Validation

Chapter 8 introduced the Cloud Computing Compliance Guideline for CSPs to be able to show compliance to customer demands regarding data location. As described in chapter 3, the validation of this guideline is carried out using interviews with a number of CSPs. This chapter describes the interview approach and validation results.

9.1 Interview approach

Interviews with CSPs are used to validate to proposed Cloud Computing Compliance Guideline. To ensure enough quality in the interviews, the CSPs to be interviewed should be able to possible future users of the guideline, meaning that they have to meet a number of criteria. Despite the fact that the focus of this research is on SaaS, SaaS providers often rely on IaaS or PaaS providers or technology, so these providers are also included. In addition, the guideline is developed for CSPs with multiple data centers in multiple countries, as data location compliance is only an issue when data can move between multiple data centers. Therefore, the selected CSPs should have multiple data centers, preferably connected to each other to exchange data, and preferable in multiple countries.

This resulted in the following criteria for the selection of CSPs:

- The CSP should use multiple data centers
- The CSP should offer IaaS, PaaS and/or SaaS services
- The CSP should preferably be active in multiple countries

Based on these criteria, a shortlist of CSPs was made, and those were approached for interviews. This resulted in five interviews, with the following companies:

- Previder
- Terremark
- Exact
- Bitbrains
- Topicus

Descriptions of these companies are available in Appendix D.

The interviews were held in the period of April 2011 – June 2011. The interview questions can be found in Appendix F. Each interview resulted in an interview report, which is approved by the interviewee.

In addition to validation of the Cloud Computing Compliance Guideline, the interviews were also used to verify the earlier gathered information about the technical infrastructure of CSPs and customer demands in cloud computing. Topics that were discussed are: general information about the CSP, customer demands, making agreements, CSP technical infrastructure, security measures, data location issues and the validation of the guideline.

9.2 Interview results

This section discusses the results of the interviews to validate the guideline, summarized per topic.

Please note that this section only discusses the results of interviews anonymously. For confidentiality reasons, the complete interview reports are only available on request.

9.2.1 Cloud Computing Compliance Guideline: general overview

The CSPs recognize the four phases of the guideline in their own processes, but do often not have it defined this strictly. It is common that a customer has certain demands, there is negotiation about the demand and supply, the service is established and configured, and sometimes reporting takes places. However, none of the CSPs uses automated negotiation.

The scope of the guideline is on a contract level. That means that agreements are made for one application or environment, and there are no agreements on e.g. a file level (e.g. allowed data locations are different per file). The CSPs agree with this level of abstraction, a smaller scope on file level is not desired, because this may need a lot of extra overhead, as the enforcement should be checked each time a file operation is carried out. In this current level of abstraction, this extra overhead only holds for a complete virtual machine.

9.2.2 Phase 1: Data location

In contraction to the cloud market experts, the CSPs mention that data location is often not an issue that customers bring up during the process. When it is the case, customers do not address it because of compliance to legislation, but because of other security demands, like the employee screening, physical access control and other aspects mentioned in ISO 27001. In the case data location is an issue, most data centers take up a contractual note that their data centers are only located within the Netherlands. This gives enough assurance for the customers. The different viewpoints of the CSPs and the cloud market experts may be because of the market segment the CSPs are in, or the fact that they state what customers do *really* ask, and the cloud markets state what customers *should* actually ask.

The CSPs mention that data processing and storage always take place within the same data center. The main reason is performance, this decreases significantly when data is processed and stored in different data centers. However, for backup, disaster recovery, redundancy etc., some CSPs store data in a different data center. However, this data center is always located within the same country.

9.2.3 Phase 2: Negotiation and agreements

Only one CSP delivers a standardized service that can is offered in a large volume, so the negotiation phase is not needed. The other CSPs attach importance to manual intake of new customers, as they deliver such specific services that they need to support the client in this process. However, the CSPs do recognize the need for standardization of the negotiation process; a formalization of terms and definitions would help to simplify the demand and supply negotiations.

The level of SLAs made differs per interviewed CSP. One CSP does not make a SLA at all, two CSPs deliver a standardized SLA that cannot be adapted, one CSP delivers a SLA that can be limitedly adapted, and one CSP negotiates the complete SLA with the customer.

One CSP mentions that it makes agreements with the customer on three levels:

- contract, which describes which services will be provided, which certifications are needed and other minimum security measures which should be taken

- SLA, which describes minimal measures that should be provided: response times, availability, throughput etc. SLAs also describe which roles are responsible for specific processes, and when should be escalated.
- "Dossier Afspraken en Procedures" (DAP), describe in natural language the processes that should be followed. These documents are created together with the client, and contain e.g. authorization matrixes, with responsible employees and phone numbers.

One CSP asks whether standards for risk assessment, data classification, security demands or negotiation (definitions of terms etc.) do already exist. These standards would facilitate the negotiation process. The SLA@SOI project is an example of standardizing these aspects.

9.2.4 Phase 2 / 3: Enforcing

A CSP suggest to make the last process step of the negotiation phase, enabling and configuring the customer environment, more explicit. This is one of the important process steps in the guideline, but is only mentioned briefly. It is suggested to make a separate phase for this step, which may replace phase 3, as this does not does not describe special process steps. The new phase will consist of the process steps 'converting agreements to configurations', 'configuring the customer environment' and 'monitoring and logging'. The conversion of agreements to agreements can be done by using the SLA@SOI framework or using the XACML framework that needs to have specific instruction for the CSP environment. The configuration of the customer environment is the actual setting of these configurations and enable of the customer environment. The last process step, monitoring and logging, are implemented to verify that the configuration still enforce the agreements, and to detect violations.

The CSP also asks the question how the physical hardware can be mapped to its geographical location. The company uses a CMDB, but naming is of the locations is not always clear enough to determine the real physical location of the hardware. It is advised to take clear naming into account when specifying the CMDB.

9.2.5 Phase 4: Reporting

At the moment, only two CSPs regularly report the results of the service levels to their clients using a Service Level Report (SLR). Two other CSPs can deliver such a SLR on report. These SLRs can include audit results and certifications, but currently do not mention data location.

An important part of the guideline is an external audit to give assurance that agreements concerning data location are enforced correctly. All CSPs would allow such an external audit. However, there should be good reasons for such an audit, because an audit takes time for the CSP and gives deep insight in the infrastructure. In addition, auditors sometimes require scripts to be executed to gather information; executing these scripts may influence the configuration or performance of other customer environments. However, CSPs argue that regular audits on data location are a good reason to allow such an audit.

9.2.6 Phase 4: Showing compliance

The guideline states that compliance can be shown by giving assurance by an external auditor. The CSPs agree with this statement. An external audit by a third party seems to be the best way to be able to show compliance. CSPs do not have any other suggestions for showing compliance.

Besides showing compliance to data location, most CSPs (or their suppliers) also show compliance to certifications like ISO 27001, SAS70. However, these certifications do not take into account data location.

9.2.7 Cloud Computing Compliance Guideline: feasibility of implementation

The CSPs think implementing the guideline into their process is feasible, at least on a high level. Some part are already in place at some CSPs, some of these parts are standard in e.g. the hypervisor, other parts are developed by the CSP itself. The CSPs consider the guideline to help them in showing compliance to their customer's demands regarding data location.

9.2.8 External validation

For the external validation, some scenarios were sketched in which the environment changes. CSPs were asked for their opinion about the guideline in the changed environment.

In the first scenario, the legislation changes. For example, customers do not longer have to store data within a specific region. The CSPs think the guideline will still hold in such a changed situation, because it generally describes the process steps to be taken. In case data location is not an issue anymore, it can be relatively easily changed to other security issues. However, some implementation details need to be made.

A second scenario is when the CSP would offer worldwide services, and connect data centers to take full advantage of the cloud opportunities. In this case, data could easily travel over the different data centers and geographical areas. The CSPs indicate that currently data does not travel between the data centers in different countries. When this will be the case, enforcement of data location needs to be stricter configured. This may also have an impact on costs; when e.g. data may be only located in the EU, but cheap energy and storage is available in the US, this should be taken into account when negotiating on the SLA.

The third scenario is about changing cloud technology. Currently, CSPs do have good control about their environment. It is possible that cloud virtualization technology becomes more advanced, and CSPs are not able to control all settings anymore, e.g. because of obfuscation[3] of the data. It is difficult for CSPs to indicate what will happen in this case, with obfuscation it is the question who is responsible for the data, and whether data is still sensitive when it is obfuscated, so it is unknown whether the guideline is still applicable. Maybe CSPs would even not adopt such technology.

9.2.9 What is missing?

The last question that was asked during the interviews was whether the CSPs were missing something in the guideline.

One CSP suggested to make the guideline a commercial product offering. Being compliant may have a competitive advantage, so it may have an added value for customers to be able to execute this

[3] Obfuscation is a technique for cutting data into smaller pieces, which are spread over multiple locations. This makes It impossible to read or steal data from one data center, as information from multiple locations is needed. Obfuscation can be used to bypass legislation, because data only becomes useful when the pieces are brought back together.

process, so they may want to spend extra money to do this. It is worth considering this option, but this is up to the CSPs themselves.

An important point mentioned by multiple CSPs is whether the scope of the guideline also includes other versions of the data for backups, disaster recovery, redundancy and mirrors. This is currently limitedly taken into account, but an important aspect to show compliance. The guideline currently only highlights the chain of suppliers, but does not explicitly mention how to address agreements with suppliers. To overcome this issue, the guideline should be with adding agreements with the suppliers of these secondary data storage providers, and adding advanced monitoring and logging for secondary data storage.

One CSP mentioned that authorization of users is not included in the guideline. This is on purpose, as it does not have a primary relation with showing compliance to data location.

Another CSP mentioned that the guideline should contain an iterative cycle: when the CSP shows compliance, the customer should check whether he still is satisfied with the offered compliance. Questions like: 'are the risks / regulations changed?' are relevant to ask. It may be needed to change the requirements after a cycle. This keeps both parties aware of the current level of compliance, and stimulates acting on this.

Some CSPs expected explicit separation between the location of storage and processing. Currently, this is not often the case, but in the future, this may be a necessary addition to the guideline. Furthermore, It is the question whether transmission of data should be included, but there are no signs that transmission of data is currently important to be able to show compliance, as e.g. the EU Data Protection Directive only mentions data storage and processing.

Some CSPs ask the question on which aspects an audit should be carried out. The model should elaborate more on this.

9.3 Conclusions

This chapter discussed the validation of the Cloud Computing Compliance Guideline, using interviews with five CSPs. The interviews brought up a number of topics that the CSPs consider as added value, and a number of topics that need improvement.

Points for improvement are the level of detail of the guideline, CSPs expect some more hand-on practical guideline and details. In addition, the guideline is limited in scope, it does not include backups and other off-site copies of the data. These aspects could be included using additional processes for governance of these 'non-primary' data, which is stored at secondary location (which may be internal or external for the CSP). This can include a process for monitoring and logging the location of such data.

The most mentioned added value is that the guideline gives an overall view of the complete process of showing compliance. The CSPs appreciate a strictly defined process, which helps them to show compliance. The CSPs think that the guideline is practically implementable on a high level, but the instructions need to be more detailed.

Based on the interviews, the guideline has been adapted slightly: the iterative circle has been added, some more information about enforcement at suppliers has been added and more detail about the

audits of the CMDB has been added. However, it was not possible to process all comments. Using the results of the interviews, the guideline has been validated. However, more validation in practice is needed to show complete validity, by e.g. using the guideline to develop a proof of concept or prototype of a technical setup at a CSP and eventually using the guideline in practice by a CSP.

10 Conclusions, discussion and future work

This final chapter gives a conclusion to the research questions, discusses the gained results, and provides directions for future work.

10.1 Conclusions

Cloud computing leads to new opportunities and procedures for both the customers and the service providers. This research is driven by the research question how cloud service providers can show compliance to customer demands regarding data location in a public SaaS cloud. This section describes the a short summary of the answers to the research questions.

10.1.1 Customer demands

RQ1: *What are customer demands regarding data location compliance?*

Interviews with experts in the cloud market have shown that customers have to show compliance to legislation, customers demand to know the location of their data, customers demand the CSP to have security certificates, customers demand a good track record by the CSP and demand assistance when migrating to the cloud. For the latter mentioned demands, solutions are already available. However, for data location and compliance, CSPs do not offer guarantees at the moment. The focus of this research will therefore be on data location, compliance and legislation, as there are still research gaps on this topic.

The EU Data Protection Directive requires customers to store and process their data within the EU. To be able to show compliance to legislation, have to determine which security demands should be requested to a CSP. Therefore, customers carry out risk assessments on data, give data a classification and determine security demands that should be enforced by the CSP. These agreements (e.g. that data is stored within the EU) are formalized in a service level agreements.

However, interviews with CSPs show that customers often do actually not demand such compliance, because there is no strict checking on data location compliance. Government agencies like the 'College bescherming persoonsgegevens' only take action after complaints by citizens. Cloud market experts state that customers actually *should* demand compliance to data location.

10.1.2 Data location

RQ2: *What technical solutions do cloud service providers currently have?*

The technical infrastructure CSPs currently have does enable them to determine and manage the location of data. The determination of data location differs per cloud computing service model; for IaaS the data location can be determined and controlled via the hypervisor, a CMDB is used to determine the geographic location of a server. For PaaS and SaaS this depends on the specific implementation by the CSP and underlying technical infrastructure. Data movement can be tracked by using logging and monitoring tools available for the CSP.

10.1.3 Current limitations

RQ3: *What are the current limitations for CSPs to show compliance to customer demands regarding data location?*

To be able to show compliance, the customer and CSP have first to make agreements on the service to be delivered. Based on these agreements, the CSP can arrange security measures which conform to the customer demands. To be sure that these measures do really comply to the customer demands, an audit is carried out check these security measures. When this process is carried out correctly, the auditor gives assurance and the CSP can show compliance to the customer demands.

In this process, the following limitations have been identified:

LIM1. The level of agreements that can be made differs per (type of) CSP. Manual negotiation takes time, and leads to contracts and SLAs that are not machine readable. Automated negotiation and creation of machine readable SLAs is not available yet.
LIM2. CSPs do not have the tools to enforce the agreements made with the customers to control the location of the data.
LIM3. Currently, there are no standardized audit instructions to give assurance about data location in cloud computing available.
LIM4. The chain of suppliers makes it more difficult to check where customer data is located, which supplier is responsible and in control for this.

10.1.4 Making agreements

RQ4: *How to make agreements about data location between customer and CSP?*

The literature study showed two SLA negotiation frameworks that can be used for making agreements between the customer and the CSP. The first approach is a combination of the XACML specification language extended for data location, together with XACML integration and the Lin and Squicciarini framework. This components are already used in practice, but not as a combination. The second approach is the SLA@SOI project, which is focused on cloud computing and SLAs and covers all aspects of the process, but this research is still in development.

10.1.5 Enforcing agreements

RQ5: *How to enforce security policies regarding location?*

To be able to enforce data location agreements, it is required to ensure correct configuration settings for each customers, added with monitoring and logging for detection violations. In addition, the mapping of physical hardware to geographical locations using the CMDB is an important aspect.

The enforcement of agreements differs per service model. IaaS can be handled by the hypervisor, PaaS can be handled by the platform and SaaS can be handled by the application. It is important that the connections between the service models that depend on each other (e.g. SaaS depends on IaaS) are configured correctly.

10.1.6 Showing compliance

RQ6: *How can cloud service providers show compliance to customer demands regarding data location in public SaaS cloud computing?*

To answer this research questions, a Cloud Computing Compliance Guideline is proposed, based on interviews and literature. The Cloud Computing Compliance Guideline gives a process description of showing compliance which enables CSPs to show compliance to customer demands regarding data location. The Cloud Computing Compliance Guideline comprises of four phases.

Phase 1 describes how the customer prepares the movement to the cloud, by carrying out a risk assessment, data classification, creating security demands regarding data location and CSP selection. Phase 2 describes the negotiation process between the customer and CSP. The guideline describes two frameworks that can be used for the SLA negotiation: the SLA@SOI framework and the XACML framework. After the automated negotiation, the CSP takes security measures to ensure data will be stored conform the agreements. Phase 3 describes the regular storage process. Because all security measures are taken, no extra efforts are needed. However, the CSP monitors and logs the movement of data, to detect possible violations. Phase 4 describes how the CSP shows compliance to the customer demands regarding data location. This is done by regularly reporting the current status, and carrying out external audits to give assurance about the correctness of the process.

When these phases are carried out correctly, an auditor checks whether CSP executes the correct processes and data is stored on the allowed locations. If this is the case, the auditor can give assurance that the agreements with the customer are enforced, so the CSP can show compliance to the customer demands.

10.2 Reflection

This section reflects on the results of this research, by looking at the contributions, and also at the limitations of this research. This is followed by a section that discusses the context of the research questions.

10.2.1 Contributions

This research identified the process of showing compliance, and made it specific for data location issues. The research gathered knowledge about the customer demands in cloud computing and about the technical infrastructure used by CSPs. It identified the possibilities for data location detection and control. Finally, the research described how auditors can give assurance about data location, and how this relates to showing compliance.

To integrate the gathered knowledge, a guideline was developed, which helps CSPs to setup a (high level) process that enables them to show compliance to customer demands. The guideline also suggests two frameworks that can be used for a practical implementation to show compliance.

The validation with CSPs shows that CSPs think the proposed guideline is useful in practice, and recognize parts of the guideline already in their infrastructure. CSPs indicate that the added value of the framework is in the strict definition of process phases and steps, and concrete guidelines for enforcement and auditing.

10.2.2 Limitations of research

Every research has its limitations, and so has this research. First of all, the concrete examples for negotiation and enforcement were introduced in the guideline relatively late. Therefore, they could not be validated during the interviews. Known limitations of the SLA@SOI framework is that it is still in the research phase, so it is not directly practically implementable. In addition, it seems to be basically designed for the IaaS service model. PaaS and SaaS service models will be added in future research. The XACML framework is a combination of existing techniques, but is not tested or validated as a framework in practice yet.

The research question explicitly mentions the SaaS service model. However, it turned out that this it is not possible to look at this service level alone: before it is possible to make statements about the SaaS service model, the underlying service models have to be investigated first. Therefore, this research also described issues on IaaS, as this is the most elementary service model, and the basis for data location determination and control. In addition, because SaaS is relatively high in the stack of offerings, multiple variants of solutions are possible. This research tried to investigate as many as possible of these variants, but does not pretend to be complete on this topic.

Despite the fact that the result of this research is a 'guideline', it is does not provide a detailed 'hands on' manual for CSPs. The guideline describes a high level overview of the complete process, and gives directions for implementation. To be able to do give a detailed hands on approach, specific knowledge of the existing infrastructure (e.g. hypervisor, negotiation process etc.) is needed.

The information in this research is mainly based on interviews with cloud experts and CSPs. Cloud computing is a relatively new topic, so little academic research is available. Despite the fact that cloud computing is relatively new, it builds on already existing techniques. In this research, it was tried to use research on these existing techniques as much as possible. However, for (dynamic) data location, this turned out to be difficult to find techniques that can be compared to cloud computing, because data location is often not in issue in previous situations.

10.2.3 Discussion

10.2.3.1 Compliance vs. trust

This research is focused completely on compliance. However, compliance is a measure that is often required by others than the customer, and does not motivate the CSP and customer themselves. Compliance seems to not be the most important aspect that drives them for handling data carefully. The proposed guideline moves the trust issue from the CSP to the auditor, customers do have at least to trust the auditor's opinion about the data location. Trust in cloud computing may evolve over time in the future, as cloud computing becomes a commodity. It is the question how long it will take before enough trust in cloud computing is gathered to exclude the need for compliance.

Currently, CSPs often have strategically located data centers; in EU and/or US. When there is no data movement between these data centers in different geographical regions, there are no problems concerning data location as long as the customer's data is stored in the correct data center. Interviews showed that customers often trust CSPs when they just mention this in a contract. However, when cloud computing techniques evolve, and data is located e.g. at the cheapest location because of electricity, the location of data may become more important to show compliance.

10.2.3.2 Legislation

The most important driver for this research is the EU Data Protection Directive, which requires European companies to store and process privacy sensitive data within the European Union. However, due to new developments it is the question whether data location will be a point of discussion in the future.

Due to the recent developments in cloud computing, the European Commission is considering changing the EU Data Protection Directive. The outcomes of these changes are still unknown, but a hearing held with Viviane Reding (the European Commissioner in charge of the review) in March

2011 suggests that the new legal data protection framework will apply to all EU citizens regardless of where the data is collected and stored [94]. This means that EU data protection rules will not only apply for organizations based within the EU, but will also apply to all organizations which are based outside the EU and process and store data of European citizens. In this case, the location of data will not be an issue anymore from a European privacy perspective, because the privacy legislation applies everywhere in the world. It is however the question whether the EU has the possibilities to effectuate this legislation worldwide, because all non EU countries have to ratify this legislation to effectuate it worldwide.

It is however the question whether the current EU data protection directive can ensure privacy, even when data is located within the European Union. Recently, an article [95] was published in which Microsoft admits that it will allow US government agencies to access data located in the European based Microsoft cloud data centers, using the Patriot Act. A Microsoft spokesman states that Microsoft is a U.S.-headquartered company, so it has to comply with local laws, as well as any other location where one of its subsidiary companies is based.

This also holds for the EU-US Safe Harbor principles. In April 2010, German data protection authorities issued a resolution requiring extra diligence for German data exporters interacting with US Safe Harbor-certified entities, holding exporters liable for lack of diligence, to face possible sanctions, effectively calling into question the sufficiency of the Safe Harbor program to meet EU guidelines. Other nations have expressed reservations about data stored in US-based clouds falling under the jurisdiction of US laws like the Patriot Act [66].

When the EU Data Protection Directive cannot guarantee protection of privacy sensitive data on EU territory, or when the directive becomes effective globally, it is the question whether compliance to data location will still be an issue. The location of data may become less important, while the struggle on which legislation applies to the data grows. In that case, the proposed guideline has to be adapted to the actual juridical circumstances.

10.3 Future work

The Cloud Computing Compliance Guideline currently only describes a process to show compliance for the primary location of data; backups, mirrors etc. are not included. To be able to show compliance, it is important that all copies of the data are compliant. Future research is needed to investigate how compliance to all copies of the data can be accomplished. Advanced monitoring and logging will probably play an important role in this research.

A recently published paper on ‘geolocation’ [66] approaches data location from a complete different perspective. It determines the location of data based on response times of the server the data is located on. The method may seem not very accurate, but when the response time is measured from multiple nodes, the authors claim this method can achieve data location accuracy up till street level. Future research on incorporating this technique in the guideline is suggested.

The current guideline has been validated with CSPs. However, some aspects were not addressed during this validation, like the exact chain of suppliers, different SaaS-options (multi-tenant, individual instances etc.). Next to that, the guideline is not checked in practice; no proof of concepts, prototype or field-experiments were carried out. It may be worth to apply these validation techniques in future research.

The current focus of the Cloud Computing Compliance Guideline is on data location only. However, the guideline can be relatively easily extended to other security aspects like authorization, replication, backups and other security measures mentioned in ISO 27002. This may also hold for the difference between storage and processing, and transportation of data. Research may be done in how to incorporate these security measures in the compliance guideline, and which auditing need to be carried out.

This research is oriented on business informatics, and modeled the guideline from a technical perspective. It may be useful to carry out research to the same topics from another perspective. From the juridical perspective, research can be done to the legislation requiring compliance. The recent developments mentioned in section may also be taken into account. From an accountancy perspective, research can be done on the specific audits that need to be carried out, and whether these are good enough to be able to show compliance. From the psychological perspective, research may be done on reasons why customers want to know the location of their data next to compliance issues; security issues and trust probably play an important role here.

References

[1] 2011 Gartner, Inc. (2011) Hype Cycles 2010. [Online]. http://www.gartner.com/technology/research/hype-cycles/

[2] KPMG, "Orchestrating the New Paradigm," KPMG, Amstelveen, White Paper 2011.

[3] Michael Armbrust et al., "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50-58, 2010.

[4] Rolf Harms and Michael Yamartino, "The economics of the cloud," Microsoft, White paper 2010.

[5] KPMG, "From Hype to Future, KPMG's 2010 Cloud Computing Survey," KPMG Advisory N.V., Amstelveen, 2010.

[6] Mark McDonald and Dave Aron, "Leading in Times of Transition: The 2010 CIO Agenda," Gartner, Inc., Report 2010.

[7] Neal Leavitt, "Is cloud computing really ready for prime time?," *Computer*, vol. 42, no. 1, pp. 15-20, January 2009.

[8] Keven J. O'Brien. (2010, September) Cloud Computing Hits Snag in Europe. [Online]. http://www.nytimes.com/2010/09/20/technology/20cloud.html

[9] European Parliament, Council , "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," *Official Journal of the EC*, vol. L281, pp. 31-50, October 1995.

[10] Rijksoverheid. (2000, July) Wet bescherming persoonsgegevens. [Online]. http://wetten.overheid.nl/BWBR0011468/#Hoofdstuk11

[11] S.M. Artz, "Derde landen," College bescherming persoonsgegevens, Den Haag, Brochure February 2002. [Online]. http://www.cbpweb.nl/pages/ind_wetten_wbp.aspx

[12] Google Inc. (2011, April) Google Apps Service Level Agreement. [Online]. http://www.google.com/apps/intl/en/terms/sla.html

[13] Online CRM. (2011, April) The Promise and Pitfalls of Service Level Agreements. [Online]. http://www.online-crm.com/sla.htm

[14] J.P.H. Donner. (2011, April) Kamerbrief over cloud computing. [Online]. http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/kamerstukken/2011/04/20/kamerbrief-over-cloud-computing/kamerbrief-over-cloud-computing.pdf

[15] Peter Mell and Tim Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, Version 15, 2009.

[16] Amazon Web Services LLC. (2011) Amazon Elastic Compute Cloud (Amazon EC2). [Online]. http://aws.amazon.com/ec2/

[17] Terremark Worldwide. (2011) The Enterpise Cloud. [Online]. http://www.terremark.com/services/cloudcomputing/theenterprisecloud.aspx

[18] Amazon Web Services LLC. (2011) AWS Elastic Beanstalk. [Online]. http://aws.amazon.com/elasticbeanstalk/

[19] Microsoft. (2011) Windows Azure | Microsoft PaaS | Cloud Services | Application Hosting. [Online]. http://www.microsoft.com/windowsazure/

[20] Salesforce.com, inc. (2011) Application Development with the Force.com Cloud Computing Platform. [Online]. http://www.salesforce.com/platform/

[21] Google. (2011) Google App Engine. [Online]. http://code.google.com/appengine/

[22] Google. (2011) GMail - Google's approach to email. [Online]. http://mail.google.com/mail/help/intl/en/about.html

[23] Microsoft. (2011) Online Software Hosted in the Cloud - Office 365. [Online]. http://www.microsoft.com/en-us/office365/online-software.aspx

[24] Salesforce.com, inc. (2011) CRM & Cloud Computing - salesforce.com. [Online]. http://www.salesforce.com

[25] Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb, "A taxonomy and survey of cloud computing systems," in *Fifth International Joint Conference on INC, IMS and IDC*, Seoul, Korea, 2009, pp. 44-51.

[26] KPMG, "The Cloud: Changing the Business Eco System," KPMG, India, 2011.

[27] Google Inc. (2011, May) Google Apps for Government. [Online]. http://www.google.com/apps/intl/en/government/index.html

[28] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung, "The Google file system," in *ACM SIGOPS Operating Systems Review*, 2003, pp. 29-43.

[29] Piet Verschuren and Hans Doorewaard, *Het ontwerpen van een onderzoek*, derde druk ed. Utrecht, The Netherlands: Uitgeverij LEMMA BV, 2005.

[30] Roel Wieringa, "Design science as nested problem solving," in *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*,

Malvern, PA, USA, 2009, p. Article No. 8.

[31] Department of Justice. (2011) The USA PATRIOT Act: Preserving Life and Liberty. [Online]. http://www.justice.gov/archive/ll/highlights.htm

[32] Anuj Saxena, *Enterprise Contract Management*.: J. Ross Publishing, 2008.

[33] One Hundred Seventh Congres of the United States of America. (2002) Sarbanes-Oxley Act of 2002. [Online]. http://news.findlaw.com/cnn/docs/gwbush/sarbanesoxley072302.pdf

[34] 104th Congress. (1996) Health Insurance Portability and Accountability Act of 1996. [Online]. https://www.cms.gov/HIPAAGenInfo/Downloads/HIPAALaw.pdf

[35] Congress USA. (2002) Federal Information Security Management Act of 2002'. [Online]. http://csrc.nist.gov/drivers/documents/FISMA-final.pdf

[36] PCI Security Standards Council LLC. (2010) Payment Application Data Security Standard. [Online]. https://www.pcisecuritystandards.org/documents/pa-dss_v2.pdf

[37] L.B. Sauerwein and J.J. Linnemann, "Handleiding voor verwerkers van persoonsgegevens," Ministerie van Justitie, Den Haag, Wet bescherming persoonsgegevens, 2002.

[38] Siani Pearson and Andrew Charlesworth, "Accountability as a way forward for privacy protection in the cloud," in *Proceedings of the First International Conference on Cloud Computing, CloudCom 2009*, vol. 5931, Beijing, China, 2009, pp. 131-144.

[39] Mike Nolan. (2011, April) Tackling Data Security Barriers to Cloud Adoption. [Online]. http://sla-at-soi.eu/2011/04/tackling-data-security-barriers-to-cloud-adoption/

[40] ISO 3166 Maintenance agency (ISO 3166/MA). (2011, June) ISO's focal point for country codes. [Online]. http://www.iso.org/iso/country_codes

[41] Gary Stoneburner, Alice Goguen, and Alexis Feringa, "Risk Management Guide for Information Technology Systems," National Institute of Standards and Technology, Falls Church, VA, USA, Special Publication 800-30, 2002.

[42] G.W. van Blarkom and J.J. Borking, "Beveiliging van persoonsgegevens," Registratiekamer, Den Haag, Achtergrondstudies en Verkenningen 23 ISBN 90 74087 27 2, 2001.

[43] Kevin Stine, Rich Kissel, William C. Barker, Jim Fahlsing, and Jessica Gulick, "Guide for Mapping Types of Information and Information Systems to Security Categories," National Institute of Standards and Technology, NIST Special Publication 2008.

[44] National Institute of Standards and Technology, "Standards for Security Categorization of Federal Information and Information Systems," National Institute of Standards and Technology, Gaithersburg, MD, USA, Federal Information Processing Standards Publication FIPS PUB 199,

2004.

[45] ISO/IEC, "Information technology — Security techniques — Information security management systems — Requirements," ISO/IEC, International Standard ISO/IEC 27001:2005(E), 2005.

[46] VMware. (2011) VMware vSphere. [Online]. http://www.vmware.com/products/drs/features.html

[47] Microsoft. (2009) Hyper-V Server 2008 R2. [Online]. http://www.microsoft.com/hyper-v-server/en/us/default.aspx

[48] Citrix Systems, Inc. (2011) Xen Hypervisor - Leading Open Source Hypervisor for Servers. [Online]. http://www.xen.org/products/xenhyp.html

[49] Aameek Singh, Madhukar Korupolu, and Dushmanta Mohapatra, "Server-storage virtualization: Integration and load balancing in data centers," in *International Conference for High Performance Computing, Networking, Storage and Analysis*, San Jose, CA, USA, 2008, pp. 1-12.

[50] Microsoft TechNet. (2011) Overview of LUN Types. [Online]. http://technet.microsoft.com/en-us/library/cc754536.aspx

[51] Adi Oltean. (2004, December) How you can uniquely identify a LUN in a Storage Area Network. [Online]. http://blogs.msdn.com/b/adioltean/archive/2004/12/30/344588.aspx

[52] David Chappell. (2010, October) Introducing the Windows Azure Platform. [Online]. http://www.microsoft.com/windowsazure/Whitepapers/introducingwindowsazureplatform/

[53] Yanpei Chen, Vern Paxson, and Randy H. Katz, "What's new about cloud computing security," *University of California, Berkeley Report No. UCB/EECS-2010-5 January*, vol. 20, January 2010.

[54] APM Group Ltd. (2011) The Official ITIL® Website. [Online]. http://www.itil-officialsite.com/home/home.aspx

[55] ITSMF Nederland, *IT Service Management, een introductie*, 3rd ed., Jan van Bon, Georges Kemmerling, and Dick Pondman, Eds. Amersfoort, the Netherlands: van Haren publishing, 2002.

[56] Siani Pearson, "Taking Account of Privacy when Designing Cloud Computing Services," in *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing*, 2009, pp. 44-52.

[57] Rajkumar Buyyaa, Chee Shin Yeoa, Srikumar Venugopala, James Broberg, and Ivona Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599-616, 2009.

[58] P. Patel, A. Ranabahu, and A. Sheth , "Service Level Agreement in Cloud Computing," in *ACM international conference on Object oriented programming systems languages and applications*,

2009.

[59] IBM. (2008, August) SPARCLE Policy Management Workbench. [Online]. http://domino.research.ibm.com/comm/research_projects.nsf/pages/sparcle.index.html

[60] Samuel Müller. (2005, February) Regulations Expressed As Logical Models (REALM). [Online]. http://www.zurich.ibm.com/security/publications/2006/REALM-at-IRIS2006-20060217.pdf

[61] Travis D. Breaux and Annie I. Antón, "Analyzing regulatory rules for privacy and security requirements," *IEEE transactions on software engineering*, vol. 34, no. 1, pp. 5-20, January 2008.

[62] OASIS, Committee Specification. (2007, April) eContracts Version 1.0. [Online]. http://docs.oasis-open.org/legalxml-econtracts/CS01/legalxml-econtracts-specification-1.0.pdf

[63] EnCoRe project. (2010) Ensuring Consent and Revocation. [Online]. http://www.encore-project.info/

[64] Marco Casassa Mont, Siani Pearson, and Pete Bramhall, "Towards accountable management of identity and privacy: Sticky policies and enforceable tracing services," in *Database and Expert Systems Applications, 2003. Proceedings. 14th International Workshop on*, 2003, pp. 377-382.

[65] (2001) Trusted Computing Platform Alliance Main Specification v1.1. [Online]. http://www.trustedcomputinggroup.org/

[66] Zachary N. J. Peterson, Mark Gondree, and Robert Beverly, "A Position Paper on Data Sovereignty: The Importance of Geolocating Data in the Cloud," in *Proceedings of the 8th USENIX conference on Networked systems design and implementation*, Portland, OR, USA, June 2011. [Online]. http://www.usenix.org/event/hotcloud11/tech/

[67] Ronald van Langeren, "Organisatie van IT-auditing," in *Grondsalgen IT-auditing*. Den Haag, Nederland: Sdu Uitgevers bv, 2005, ch. 6, pp. 144-146.

[68] Richard Chow et al., "Controlling data in the cloud: outsourcing computation without outsourcing control," in *Proceedings of the 2009 ACM workshop on Cloud computing security*, Chicago, IL, USA, 2009, pp. 85-90.

[69] Zhixiong Chen and John Yoon, "IT Auditing to Assure a Secure Cloud Computing," in *Proceedings of the 6th World Congress on Services*, Miami, FL, USA, 2010, pp. 253-259.

[70] AICPA, "Service Organizations," AICPA, Standard SAS 70, 1992.

[71] Ronald Jonker, *IT-auditing: Third Party Mededelingen en SAS 70-onderzoeken*. Den Haag, The Netherlands: Sdu Uitgevers bv, 2007.

[72] Z. Rezaee, R. Elam, and A. Sharbatoghlie, "Continuous auditing: the audit of the future," *Managerial Auditing Journal*, vol. 16, no. 3, pp. 150-158, 2001.

[73] W3C Policy Languages Interest Group. (2009, May) Review of Policy Languages and Frameworks. [Online]. http://www.w3.org/Policy/pling/wiki/PolicyLangReview

[74] W3C. (2007, November) Platform for Privacy Preferences (P3P) Project. [Online]. http://www.w3.org/P3P/

[75] Policy Group, Department of Computing, Imperial College London, UK. Ponder: A Policy Language for Distributed Systems Management. [Online]. http://www-dse.doc.ic.ac.uk/Research/policies/ponder.shtml

[76] Policy Group, Department of Computing, Imperial College London, UK. (2011, April) Ponder2 Wiki. [Online]. http://www.ponder2.net/

[77] L3S Research Center, Germany. (2008, May) Protune - PROvisional TrUst NEgotiation. [Online]. http://policy.l3s.uni-hannover.de/

[78] UMBC ebiquity research group. (2005, May) Rei: a policy specification language. [Online]. http://rei.umbc.edu/

[79] OASIS OpenOASIS Open, "Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0 – Errata Composite," OASIS Open, Working Draft sstc-saml-core-errata-2.0-wd-06, 2009.

[80] OASIS Open, "eXtensible Access Control Markup Language (XACML) version 2.0," OASIS Open, OASIS Standard oasis-access_control-xacml-2.0-core-spec-os, 2005.

[81] Sun Microsystems, Inc. (2003, March) A Brief Introduction to XACML. [Online]. http://www.oasis-open.org/committees/download.php/2713/Brief_Introduction_to_XACML.h

[82] Open Geospatial Consortium. (2011, May) Geospatial eXtensible Access Control Markup Language (GeoXACML). [Online]. http://www.opengeospatial.org/standards/geoxacml

[83] P. Mazzoleni, E. Bertino, B. Crispo, and S. Sivasubramanian, "XACML Policy Integration Algorithms," *ACM Transactions on Information and System Security (TISSEC)*, vol. 11, no. 1, pp. 1-29, 2008.

[84] IBM Corporation, "Web Service Level Agreement (WSLA) Language Specification," IBM Corporation, Language specification wsla-2003/01/28, 2003.

[85] SLA@SOI Project. (2011) SLA@SOI: Empowering the service industry with SLA-aware infrastructures. [Online]. http://sla-at-soi.eu

[86] Dan Lin and Anna Squicciarini, "Data protection models for service provisioning in the cloud," in *Proceeding of the 15th ACM symposium on Access control models and technologies*, Pittsburg, USA, 2010, pp. 183-192.

[87] SLA@SOI Project. (2009, May) Deliverable D.A5.a. [Online]. http://sla-at-soi.eu/wp-content/uploads/2009/10/D.A5a-M12-SLA-Foundations-and-Management.pdf

[88] SLA@SOI Project. (2010, September) Public Version of Enterprise IT Use Case, Lab Demonstrator. [Online]. http://sla-at-soi.eu/wp-content/uploads/2008/12/SLA@SOI_EntIT_LabDemonstrator.pdf

[89] Dan Lin, Prathima Rao, Elisa Bertino, and Jorge Lobo, "An Approach to Evaluate Policy Similarity," in *Proceedings of the 12th ACM symposium on Access control models and technologies*, 2007, pp. 1-10.

[90] M. Fujita, P.C. McGeer, and J.C.-Y. Yang, "Multi-terminal binary desision deagrams: An efficient datastructure for matrix representation.," *Formal Methods in System Design*, vol. 10, no. 2-3, pp. 149-169, 1997.

[91] S. Cabuk et al., "Towards automated security policy enforcement in multi-tenant virtual data centers," *Journal of Computer Security*, vol. 18, no. 1, pp. 89-121, 2010.

[92] SLA @ SOI Project. (2009, December) Challenges in SLA Translation. [Online]. http://sla-at-soi.eu/wp-content/uploads/2009/12/ChallengesInSLATranslation.pdf

[93] Rob Fijneman, Edo Roos Lindgreen, and Piet Veltman, *Grondslagen IT-auditing*. Den Haag, The Netherlands: Sdu Uitgevers BV, 2005.

[94] Bobbie Johnson. (2011, March) U.S. Web Firms Told to Stick to EU Privacy Laws. [Online]. http://gigaom.com/2011/03/17/u-s-web-firms-told-to-stick-to-eu-privacy-laws/

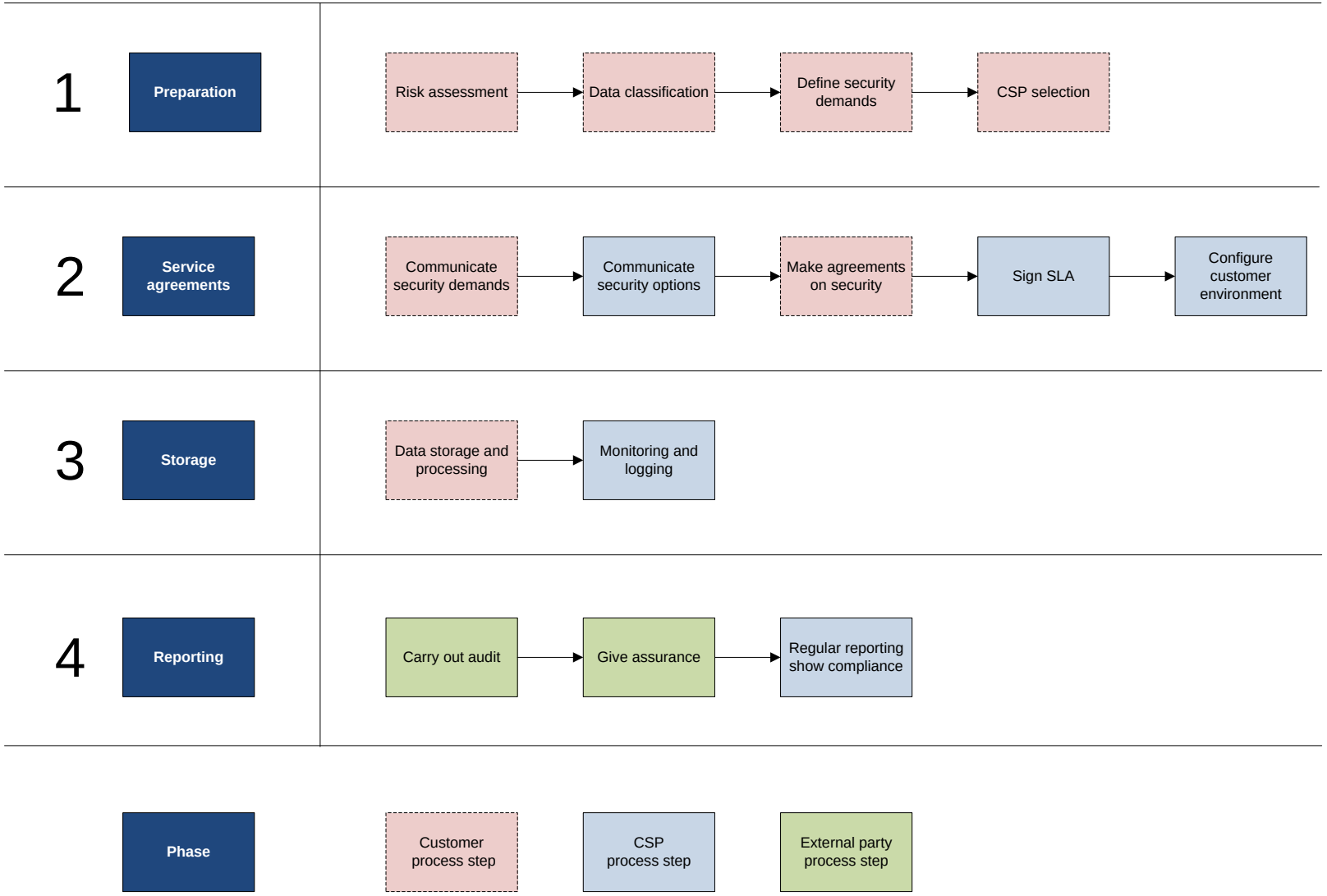
[95] Zack Whittaker. (2011, June) Microsoft admits Patriot Act can access EU-based cloud data. [Online]. http://www.zdnet.com/blog/igeneration/microsoft-admits-patriot-act-can-access-eu-based-cloud-data/11225

Abbreviations

CBP	College Bescherming Persoonsgegevens
CCCG	Cloud Computing Compliance Guideline
CIA	Confidentiality, Integrity, Availability
CMDB	Configuration Management Database
CSP	Cloud Service Provider
DAP	Dossier Afspraken en Procedures
FISMA	Federal Information Security Management Act
HDD	Hard disk drive
HIPAA	Health Insurance Portability and Accountability Act
IaaS	Infrastructure as a Service
IBE	Identifier-based Encryption
ITIL	Information Technology Infrastructure Library
KPI	Key Performance Indicator
L&S	Lin and Squicciarini
LUN	Logical Unit Number
NAS	Network Attached Storage
NIST	National Institute of Standards and Technology
PaaS	Platform as a Service
PCI DSS	Payment Card Industry Data Security Standard
PDP	Policy Decision Point
PEP	Policy Enforcement Point
SaaS	Software as a Service
SAN	Storage Area Network
SLA	Service Level Agreement
SLR	Service Level Report
SOx	Sarbanes-Oxley Act
TCPA	Trusted Computing Platform Alliance
TPM	Trusted Platform Module
VM	Virtual Machine
Wbp	Wet bescherming persoonsgegevens
WSLA	Web Service Level Agreement
XACML	eXtensible Access Control Markup Language

Appendices

Appendix A Cloud Computing Compliance Guideline



Appendix B Directive 95/46/EC of the European Parliament and of the Council

of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

CHAPTER IV TRANSFER OF PERSONAL DATA TO THIRD COUNTRIES

Article 25

Principles

1. The Member States shall provide that the transfer to a third country of personal data which are undergoing processing or are intended for processing after transfer may take place only if, without prejudice to compliance with the national provisions adopted pursuant to the other provisions of this Directive, the third country in question ensures an adequate level of protection.

2. The adequacy of the level of protection afforded by a third country shall be assessed in the light of all the circumstances surrounding a data transfer operation or set of data transfer operations; particular consideration shall be given to the nature of the data, the purpose and duration of the proposed processing operation or operations, the country of origin and country of final destination, the rules of law, both general and sectoral, in force in the third country in question and the professional rules and security measures which are complied with in that country.

3. The Member States and the Commission shall inform each other of cases where they consider that a third country does not ensure an adequate level of protection within the meaning of paragraph 2.

4. Where the Commission finds, under the procedure provided for in Article 31 (2), that a third country does not ensure an adequate level of protection within the meaning of paragraph 2 of this Article, Member States shall take the measures necessary to prevent any transfer of data of the same type to the third country in question.

5. At the appropriate time, the Commission shall enter into negotiations with a view to remedying the situation resulting from the finding made pursuant to paragraph 4.

6. The Commission may find, in accordance with the procedure referred to in Article 31 (2), that a third country ensures an adequate level of protection within the meaning of paragraph 2 of this Article, by reason of its domestic law or of the international commitments it has entered into, particularly upon conclusion of the negotiations referred to in paragraph 5, for the protection of the private lives and basic freedoms and rights of individuals.

Member States shall take the measures necessary to comply with the Commission's decision.

Article 26

Derogations

1. By way of derogation from Article 25 and save where otherwise provided by domestic law governing particular cases, Member States shall provide that a transfer or a set of transfers of

personal data to a third country which does not ensure an adequate level of protection within the meaning of Article 25 (2) may take place on condition that:

(a) the data subject has given his consent unambiguously to the proposed transfer; or

(b) the transfer is necessary for the performance of a contract between the data subject and the controller or the implementation of precontractual measures taken in response to the data subject's request; or

(c) the transfer is necessary for the conclusion or performance of a contract concluded in the interest of the data subject between the controller and a third party; or

(d) the transfer is necessary or legally required on important public interest grounds, or for the establishment, exercise or defence of legal claims; or

(e) the transfer is necessary in order to protect the vital interests of the data subject; or

(f) the transfer is made from a register which according to laws or regulations is intended to provide information to the public and which is open to consultation either by the public in general or by any person who can demonstrate legitimate interest, to the extent that the conditions laid down in law for consultation are fulfilled in the particular case.

2. Without prejudice to paragraph 1, a Member State may authorize a transfer or a set of transfers of personal data to a third country which does not ensure an adequate level of protection within the meaning of Article 25 (2), where the controller adduces adequate safeguards with respect to the protection of the privacy and fundamental rights and freedoms of individuals and as regards the exercise of the corresponding rights; such safeguards may in particular result from appropriate contractual clauses.

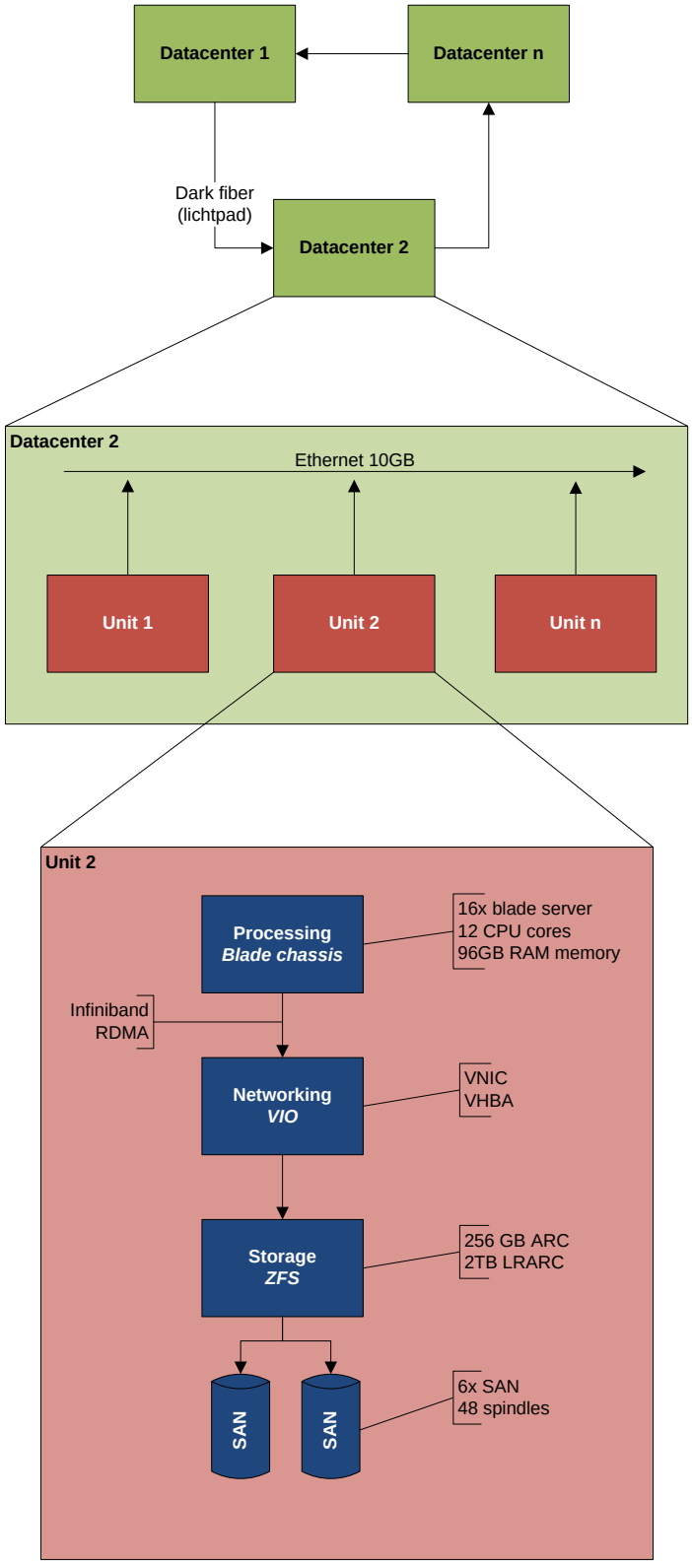
3. The Member State shall inform the Commission and the other Member States of the authorizations it grants pursuant to paragraph 2.

If a Member State or the Commission objects on justified grounds involving the protection of the privacy and fundamental rights and freedoms of individuals, the Commission shall take appropriate measures in accordance with the procedure laid down in Article 31 (2).

Member States shall take the necessary measures to comply with the Commission's decision.

4. Where the Commission decides, in accordance with the procedure referred to in Article 31 (2), that certain standard contractual clauses offer sufficient safeguards as required by paragraph 2, Member States shall take the necessary measures to comply with the Commission's decision.

Appendix C CSP cloud architecture



Appendix D Interviews

This appendix gives an overview of the interviews held during this research.

Date	Company	Interviewee	Other attendants	Topics
15-12-2010	KPMG	Koos Wolters	-	Risk assessment Data classification
03-02-2011	KPMG	Sander Klous	-	Policy exchange Data location in VM's
11-02-2011	KPMG	Mike Chung	-	Cloud audits
24-02-2011	Bitbrains	Gjalt van Rutten	Aziz Ait Ali Mark Butterhoff	Customer use cases Performance (Aziz) Data location in VM's
25-02-2011	KPMG	Serge Wallagh	Aziz Ait Ali Jonathan Chin Sue	Customer use cases Performance (Aziz) Data location
25-02-2011	KPMG	Mike Chung	Aziz Ait Ali	Customer use cases Performance (Aziz)
22-03-2011	Fleishman-Hillard	Koen Droste	-	EU privacy legislation
29-03-2011	Previder	Peter Bult	-	Cloud data center, validation model
01-04-2011	Terremark	Hans Reinhart Michiael van Til	Aziz Ait Ali Tunde Balint	Cloud management interface
05-04-2011	Terremark	Hans Reinhart	Aziz Ait Ali	Cloud proposition, validation model
23-05-2011	Exact Online	Timo van Noppen	-	SaaS, validation model
24-05-2011	Bitbrains	Gjalt van Rutten	Aziz Ait Ali	Performance (Aziz), validation model
15-06-2011	Equinix	Mark Hurd Derek Jager	Mike Chung	Data center strategy
21-06-2011	Topicus	Mario Peters	-	SaaS data centers, validation model
21-06-2011	Topicus	Marteniek Bierman	-	Cloud computing governance, validation model

Fleishman-Hillard

Fleishman-Hillard Inc., is a strategic communications firms. Based in St. Louis, the firm operates throughout North America, Europe, Asia Pacific, Middle East, Africa and Latin America through its 80 owned offices. Fleishman-Hillard is a part of Omnicom Group Inc, a global advertising, marketing and corporate communications company. Omnicom's branded networks and specialty firms provide advertising, strategic media planning and buying, interactive, direct and promotional marketing, public relations and other specialty communications services to more than 5,000 clients in more than 100 countries.

Bitbrains

Bitbrains is a cloud service provider (IaaS, PaaS) that offers Managed Hosting (a virtual private data center). Bitbrains is the first vCloud (VMware) vendor in EU. The customers of Bitbrains are financial enterprises with business critical and High Performance Computing systems. These systems require efficient, secure, scalable, and high performance infrastructure.

Previder

Previder (formerly known as Introweb) is a cloud service provider located in Hengelo (OV). Previder uses a partner network to offer its services. This means that only Previder partners offer the services to the end consumers. These end customers include web hosters, Independent Software Vendors (ISVs) and SaaS providers.

Terremark

Terremark is a provider of information technology services. With data centers in the United States, Europe and Latin America and access to massive and diverse network connectivity, Terremark delivers services which include managed hosting, colocation, disaster recovery, security, data storage and cloud computing services. Recently, Verizon bought Terremark to expand its cloud services.

Exact

Exact develops since 1984 ERP, CRM, HRM, financial software, accounting software en business software for wholesale, production, service providers en accountants. Since 2005, Exact offers accounting software online via the Software as a Service model (Exact Online). This software is targeted at the small companies (less than 20 employees) and their accountants.

Topicus

Topicus is a ICT service provider, with about 270 employees, located in Deventer, Zwolle and Enschede. Topicus is specialized in chain integration, the realization of SaaS (Software as a Service) applications and en process management for the sectors finance, healthcare and education.

Topicus is organized in cells, which all are separate units servicing different customer segments. The company has clients in three areas: education, finance and healthcare. Specific security measures are most demanded by financials, the governance aspect is important for the healthcare domain.

Appendix E Cloud expert interview questions

The interviews have the goal to get an overview of the demands customers have in cloud computing, how these demands differ from traditional IT, why customers have these demands, and how customers expect them to be fulfilled by CSPs. The results of these interviews are used to determine the current limitations customers experience concerning their demands.

The following interview questions will guide the interview:

1. There are different types of users for cloud services. Which cloud user segments can be distinguished? (customers, B2B, small/midsize, enterprises)
   - Which segments are likely to move to the cloud, which not?
2. Which propositions do you offer to clients, what is the business model?
   - Which service model? (IaaS, PaaS, SaaS)
   - Which deployment model? (private cloud, public cloud, hybrid cloud, community cloud)
3. What are client demands (use cases) which we cannot address yet?
4. Some types of applications / services are fully ready to be transferred to the cloud, others can only be used on-premise. What are examples of the different cloud-readiness phases?
5. What are barriers for customers to migrate applications to the cloud?
   - Is knowing the location of their data an important requirement / barrier for customers to go into the cloud?
6. Which additional requirements do customers have for cloud applications?

Appendix F CSP Interview questions

The goal of the CSP interviews is twofold. The first goal is to get an overview of the current technical infrastructure and relation with the customers. The second goal is to validate the Cloud Computing Compliance Guideline.

General information about the company

Company & customers

- About the offered services
- What are typical use cases for cloud computing?

Client demands

- Do Dutch and **international** clients have different demands?
- Do clients **demand** guarantees about the **location** of their data in the cloud?
- Does the CSP offer **standardized SLAs**, or do clients have their own, personalized SLAs?

Technical background

Data storage process

Verification of conceptualization of the data storage process

- How is data **stored** in VMs? (LUN per clients)
- Is **data location logged** / traceable? (what can be found in logs, what is technically possible)
- Is it possible / desired for clients to make **dynamic agreements** on the service delivered?
- To which degree are clients in **control** of the cloud **settings**, and what has to be controlled by he CSP?
- Are hardware components tracked within a **Component Management Database** (CMDB)?

Security measures

- Is the CSP ISO 27002 / SAS70 / FISMA / … **compliant**? On which areas?
- Do clients **demand compliance** to these standards?
- Is **encryption** used? (enforced / optional, on which level)
- On which level can **customer demand policies** be defined? (LUN level, database level, …) (note difference for SaaS vs IaaS / PaaS providers)
- Are some activities outsourced to **subcontractors**? What is the **impact** for security?

Validation hypotheses

Reporting

- Which service levels are **reported** to the customer regularly? (Service Level Reporting, what about security reporting, incidents)
- Would it be desired to report aspects like continuous auditing, location logs / monitoring, …?
- How does service level reporting take place at the moment?
  - Is security, location etc. included?

Validation model

Compliance in Cloud Computing

- Check attached **Cloud Computing Compliance Guideline**
  - Check data flow model
  - Check data center **infrastructure** model
  - Check conceptual model
- Is data always **stored** on the location where the data is **processed**? (assumption in model)
- Discuss concrete **scenarios** like data movement because of e.g. cheap electricity
- Discuss the impact of **subcontractors**, data location guarantees, processing / exchanging policies
- What is the **impact** of implementing enforcement of policies? How could this be done?
- Is it desirable to exchange policies via **API's**?
- Would the CSP allow an **audit** by external auditors?
- What is **missing** in the model that should be included?
- Would the CSP allow **hardware tokens** to their machine park to guarantee the location?
- Is the CSP familiar with **SLA@SOI** and/or **XACML** policies?

External validation

- What would happen with the guideline when legislation changes? E.g. it is no longer required to store data within specific geographic regions?
- What would happen with the guideline when the CSP would offer worldwide services, and interconnects data centers to easy move data around the world?
- What would happen with the guideline when cloud technology changes? E.g. CSPs would have less control over data location with obfuscation techniques.
- What would happen with the model when customer demands change? E.g. when customers have different demands regarding data location / legislation.